Chapter 1 Solutions

1.1. Most students will prefer to work in seconds, to avoid having to work with decimals or fractions.

1.2. Who? The individuals in the data set are students in a statistics class. **What?** There are eight variables: ID (a label, with no units); Exam1, Exam2, Homework, Final, and Project (in units in "points," scaled from 0 to 100); TotalPoints (in points, computed from the other scores, on a scale of 0 to 900); and Grade (A, B, C, D, and E). **Why?** The primary purpose of the data is to assign grades to the students in this class, and (presumably) the variables are appropriate for this purpose. (The data might also be useful for other purposes.)

1.3. Exam1 = 79, Exam2 = 88, Final = 88.

1.4. For this student, TotalPoints $= 2 \cdot 86 + 2 \cdot 82 + 3 \cdot 77 + 2 \cdot 90 + 80 = 827$, so the grade is B.

1.5. The cases are apartments. There are five variables: rent (quantitative), cable (categorical), pets (categorical), bedrooms (quantitative), distance to campus (quantitative).

1.6. (a) To find injuries per worker, divide the rates in Example 1.6 by 100,000 (or, redo the computations *without* multiplying by 100,000). For wage and salary workers, there are 0.000034 fatal injuries per worker. For self-employed workers, there are 0.000099 fatal injuries per worker. **(b)** These rates are 1/10 the size of those in Example 1.6, or 10,000 times larger than those in part (a): 0.34 fatal injuries per 10,000 wage/salary workers, and 0.99 fatal injuries per 10,000 self-employed workers. **(c)** The rates in Example 1.6 would probably be more easily understood by most people, because numbers like 3.4 and 9.9 feel more "familiar." (It might be even better to give rates per *million* worker: 34 and 99.)

1.7. Shown are two possible stemplots; the first uses split stems (described on page 11 of the text). The scores are slightly left-skewed; most range from 70 to the low 90s.

5	58
6	0
6	58
7	0023
7	5558
8	00003
8	5557
9	0002233
9	8

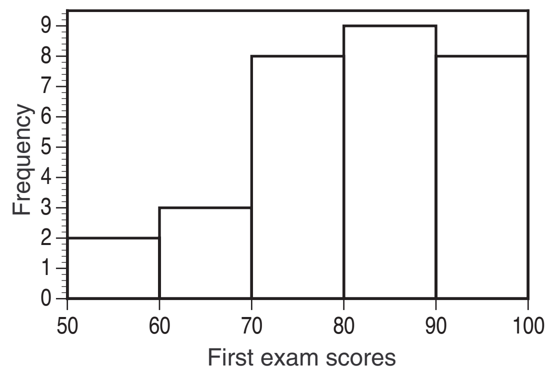
5	58
6	058
7	00235558
8	000035557
9	00022338

1.8. Preferences will vary. However, the stemplot in Figure 1.8 shows a bit more detail, which is useful for comparing the two distributions.

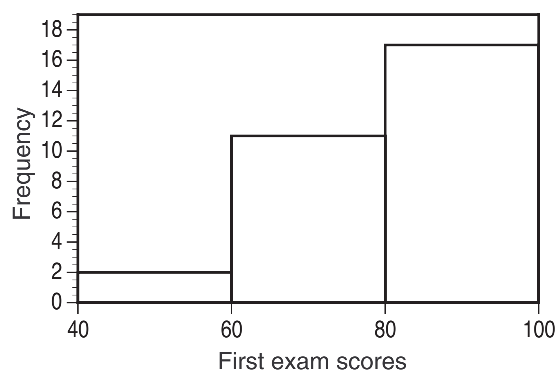
1.9. (a) The stemplot of the altered data is shown on the right. **(b)** Blank stems should always be retained (except at the beginning or end of the stemplot), because the gap in the distribution is an important piece of information about the data.

1	6
2	
2	5568
3	34
3	55678
4	012233
4	8
5	1

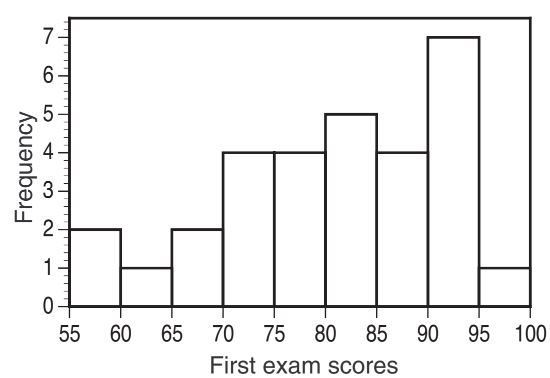
1.10. Student preferences will vary. The stemplot has the advantage of showing each individual score. Note that this histogram has the same shape as the second histogram in Exercise 1.7.



1.11. Student preferences may vary, but the larger classes in this histogram hide a lot of detail.



1.12. This histogram shows more details about the distribution (perhaps more detail than is useful). Note that this histogram has the same shape as the first histogram in the solution to Exercise 1.7.



1.13. Using either a stemplot or histogram, we see that the distribution is left-skewed, centered near 80, and spread from 55 to 98. (Of course, a histogram would not show the exact values of the maximum and minimum.)

1.14. **(a)** The cases are the individual employees. **(b)** The first four (employee identification number, last name, first name, and middle initial) are labels. Department and education level are categorical variables; number of years with the company, salary, and age are quantitative variables. **(c)** Column headings in student spreadsheets will vary, as will sample cases.

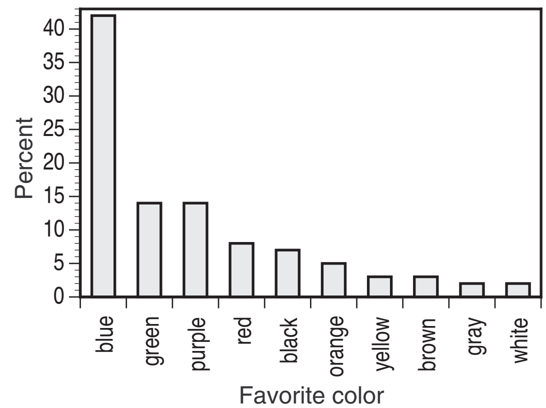
1.15. A Web search for "city rankings" or "best cities" will yield lots of ideas, such as crime rates, income, cost of living, entertainment and cultural activities, taxes, climate, and school system quality. (Students should be encouraged to think carefully about how some of these might be quantitatively measured.)

1.16. Recall that categorical variables place individuals into groups or categories, while quantitative variables "take numerical values for which arithmetic operations... make sense."

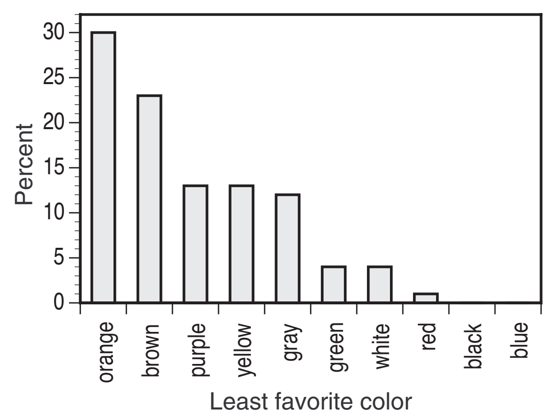
Variables (a), (d), and (e)—age, amount spent on food, and height—are quantitative. The answers to the other three questions—about dancing, musical instruments, and broccoli—are categorical variables.

1.18. Student answers will vary. A Web search for "college ranking methodology" gives some ideas; in recent year, *U.S. News and World Report* used "16 measures of academic excellence," including academic reputation (measured by surveying college and university administrators), retention rate, graduation rate, class sizes, faculty salaries, student-faculty ratio, percentage of faculty with highest degree in their fields, quality of entering students (ACT/SAT scores, high school class rank, enrollment-to-admission ratio), financial resources, and the percentage of alumni who give to the school.

1.19. For example, blue is by far the most popular choice; 70% of respondents chose 3 of the 10 options (blue, green, and purple).



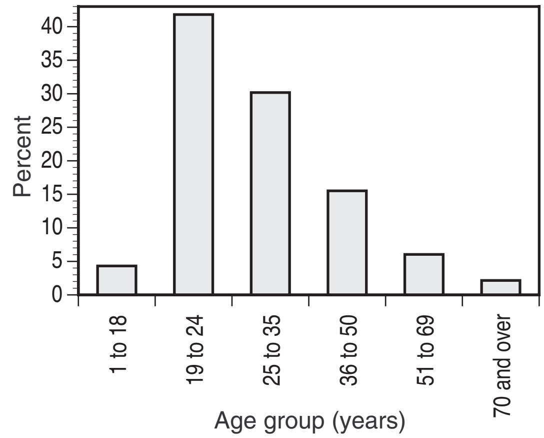
1.20. For example, opinions about least-favorite color are somewhat more varied than favorite colors. Interestingly, purple is liked and disliked by about the same fractions of people.



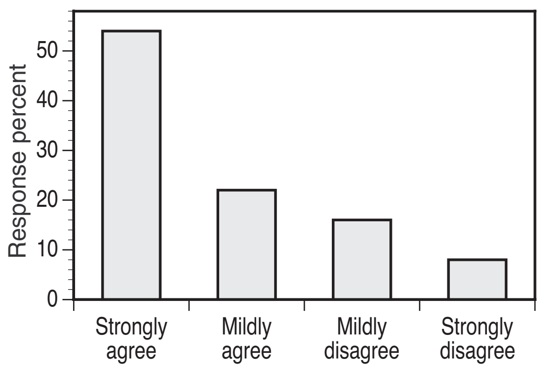
1.21. **(a)** There were 232 total respondents. The table that follows gives the percents; for example, $\frac{10}{232} \doteq 4.31\%$. **(b)** The bar graph is on the following page. **(c)** For example, 87.5% of the group were between 19 and 50. **(d)** The age-group classes do not have equal width: The first is 18 years wide, the second is 6 years wide, the third is 11 years wide, etc.

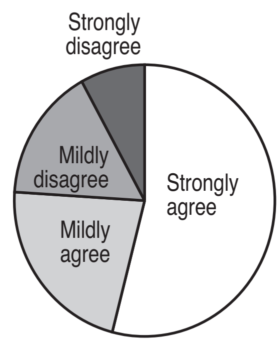
Note: *In order to produce a histogram from the given data, the bar for the first age group would have to be three times as wide as the second bar, the third bar would have to be wider than the second bar by a factor of 11/6, etc. Additionally, if we change a bar's* *width by a factor of x, we would need to change that bar's height by a factor of 1/x.*

Age group (years)	Percent
1 to 18	4.31%
19 to 24	41.81%
25 to 35	30.17%
36 to 50	15.52%
51 to 69	6.03%
70 and over	2.16%

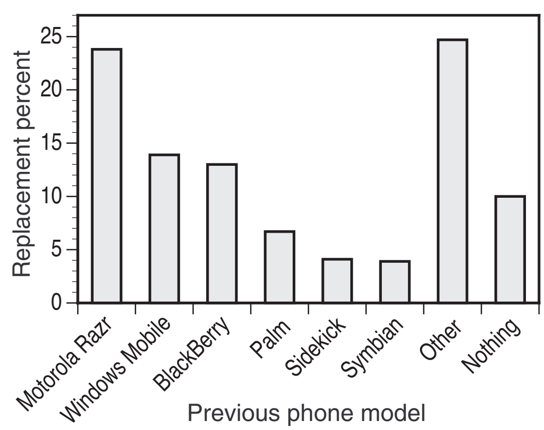


1.22. **(a)** & **(b)** The bar graph and pie charts are shown below. **(c)** A clear majority (76%) agree or strongly agree that they browse more with the iPhone than with their previous phone. **(d)** Student preferences will vary. Some might prefer the pie chart because it is more familiar.

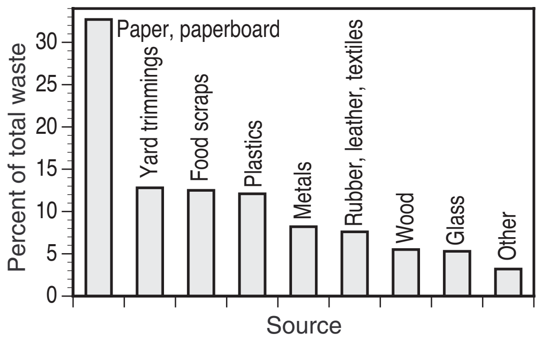


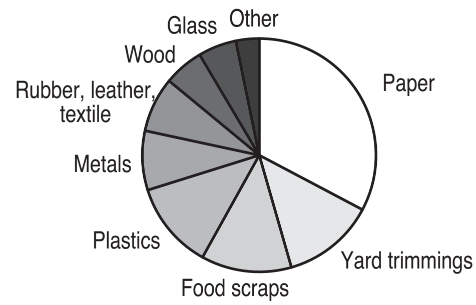


1.23. Ordering bars by decreasing height shows the models most affected by iPhone sales. However, because "other phone" and "replaced nothing" are different than the other categories, it makes sense to place those two bars last (in any order).

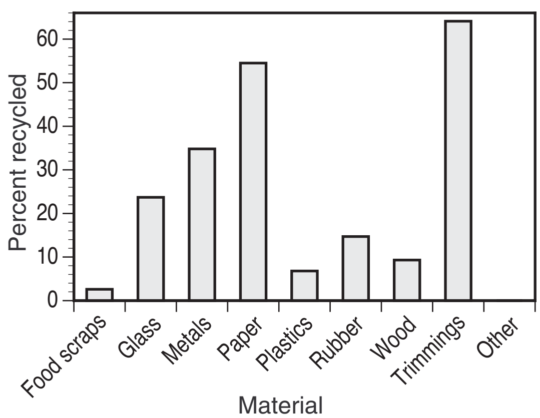


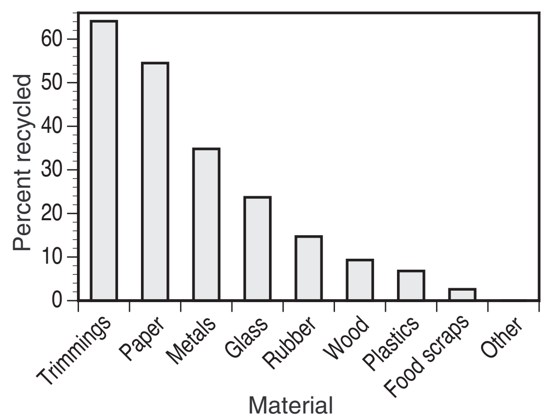
1.24. **(a)** The weights add to 254.2 million tons, and the percents add to 99.9.
(b) & **(c)** The bar graph and pie chart are shown below.



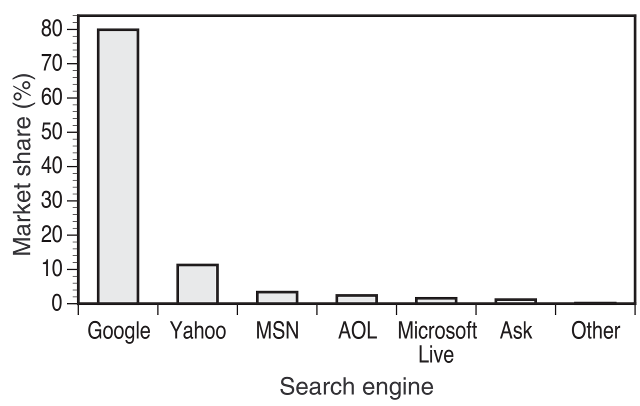


1.25. **(a)** & **(b)** Both bar graphs are shown below. **(c)** The ordered bars in the graph from (b) make it easier to identify those materials that are frequently recycled and those that are not. **(d)** Each percent represents part of a different whole. (For example, 2.6% of *food scraps* are recycled; 23.7% of *glass* is recycled, etc.)

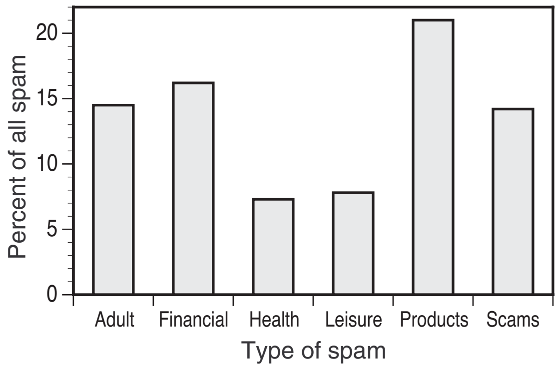


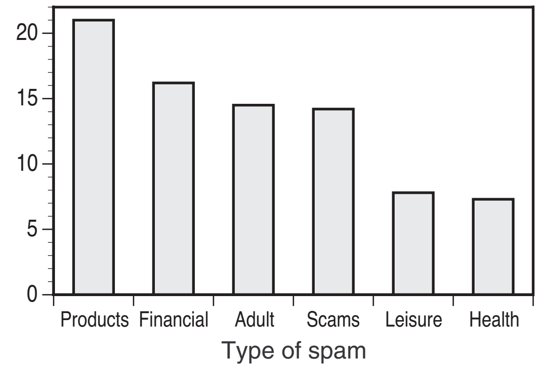


1.26. **(a)** The bar graph is shown on the right. **(b)** The graph clearly illustrates the dominance of Google; its bar dwarfs those of the other search engines.

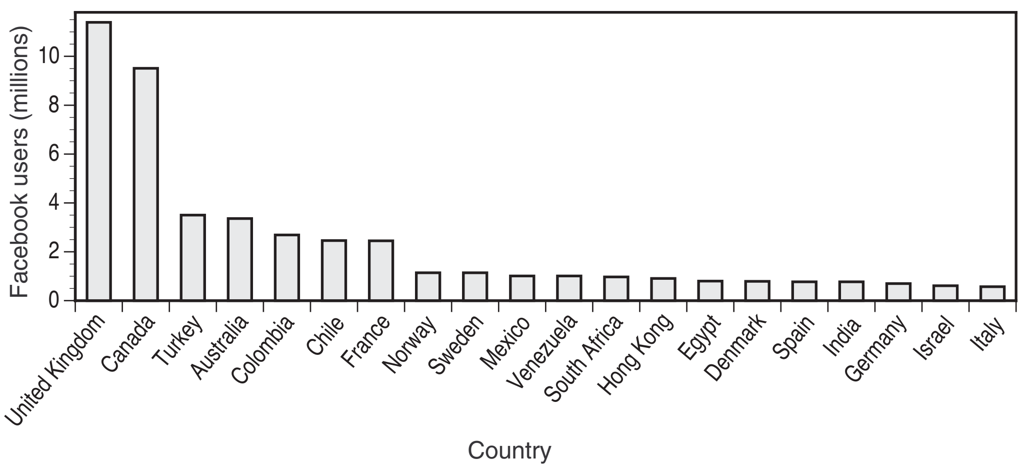


1.27. The two bar graphs are shown below.





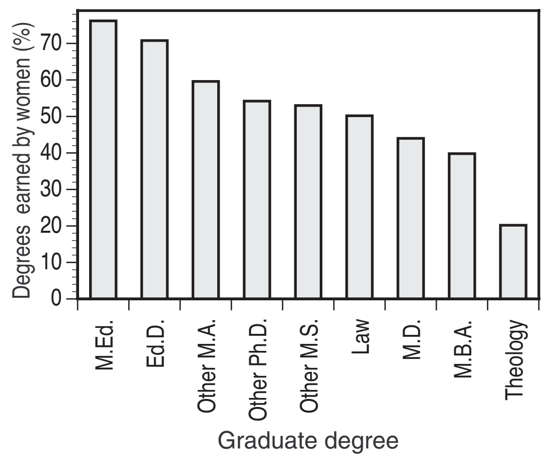
1.28. **(a)** The bar graph is below. **(b)** The number of Facebook users trails off rapidly after the top seven or so. (Of course, this is due in part to the variation in the populations of these countries. For example, that Norway has nearly half as many Facebook users as France is remarkable, because the 2008 populations of France and Norway were about 62.3 million and 4.8 million, respectively.)



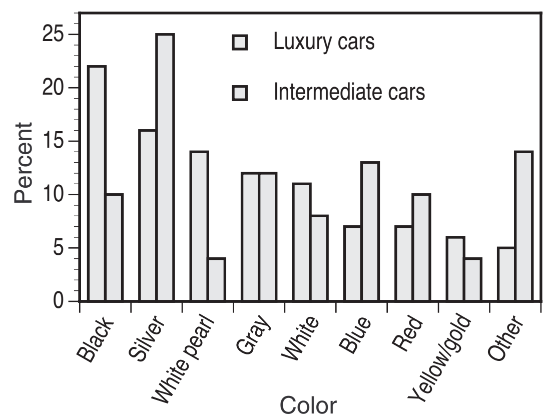
1.29. **(a)** Most countries had moderate (single- or double-digit) increases in Facebook usages. Chile (2197%) is an extreme outlier, as are (maybe) Venezuela (683%) and Colombia (246%). **(b)** In the stemplot on the right, Chile and Venezuela have been omitted, and stems are split five ways. **(c)** One observation is that, even without the outliers, the distribution is right-skewed. **(d)** The stemplot can show some of the detail of the low part of the distribution, if the outliers are omitted.

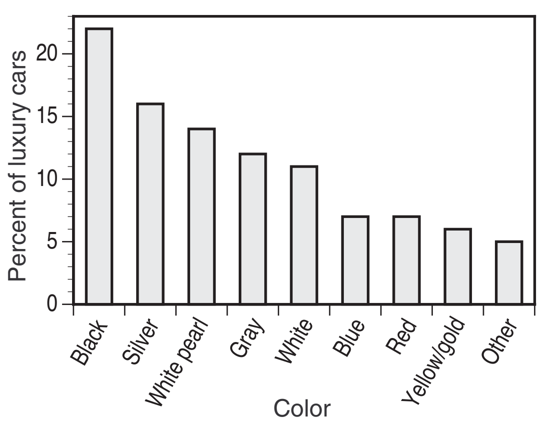
0	000
0	2333
0	4444
0	6
0	99
1	
1	33
1	
1	
1	
2	
2	
2	4

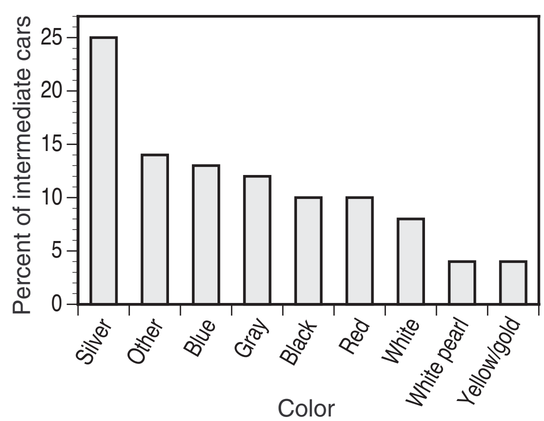
1.30. **(a)** The given percentages refer to nine distinct groups (all M.B.A. degrees, all M.Ed. degrees, and so on) rather than one single group. **(b)** Bar graph shown on the right. Bars are ordered by height, as suggested by the text; students may forget to do this or might arrange in the opposite order (smallest to largest).



1.31. **(a)** The luxury car bar graph is below on the left; bars are in decreasing order of size (the order given in the table). **(b)** The intermediate car bar graph is below on the right. For this stand-alone graph, it seemed appropriate to re-order the bars by decreasing size. Students may leave the bars in the order given in the table; this (admittedly) might make comparison of the two graphs simpler. **(c)** The graph on the right is one possible choice for comparing the two types of cars: for each color, we have one bar for each car type.







1.32. This distribution is skewed to the right, meaning that Shakespeare's plays contain many short words (up to six letters) and fewer very long words. We would probably expect most authors to have skewed distributions, although the exact shape and spread will vary.

1.33. Shown is the stemplot; as the text suggests, we have trimmed numbers (dropped the last digit) and split stems. 359 mg/dl appears to be an outlier. Overall, glucose levels are not under control: Only 4 of the 18 had levels in the desired range.

0	799
1	0134444
1	5577
2	0
2	57
3	
3	5

1.34. The back-to-back stemplot on the right suggests that the individual-instruction group was more consistent (their numbers have less spread) but not more successful (only two had numbers in the desired range).

Individual		Class
	0	799
22	1	0134444
99866655	1	5577
22222	2	0
8	2	57
	3	
	3	5

1.35. The distribution is roughly symmetric, centered near 7 (or "between 6 and 7"), and spread from 2 to 13.

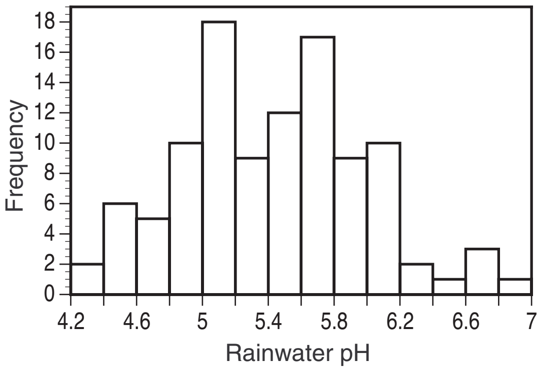
1.36. **(a)** Totals emissions would almost certainly be higher for very large countries; for example, we would expect that even with great attempts to control emissions, China (with over 1 billion people) would have higher total emissions than the smallest countries in the data set. **(b)** A stemplot is shown; a histogram would also be appropriate. We see a strong right skew with a peak from 0 to 0.2 metric tons per person and a smaller peak from 0.8 to 1. The three highest countries (the United States, Canada, and Australia) appear to be outliers; apart from those countries, the distribution is spread from 0 to 11 metric tons per person.

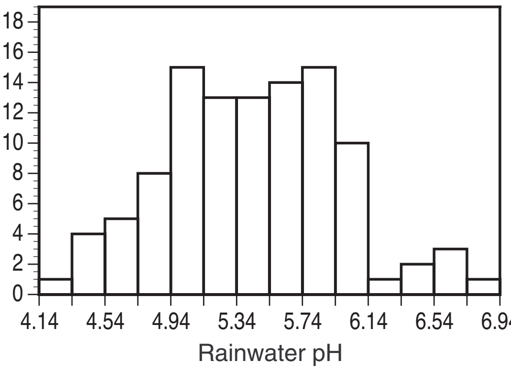
0	00000000000000011111
0	222233333
0	445
0	6677
0	888999
1	001
1	
1	
1	67
1	9

1.37. To display the distribution, use either a stemplot or a histogram. DT scores are skewed to the right, centered near 5 or 6, spread from 0 to 18. There are no outliers. We might also note that only 11 of these 264 women (about 4%) scored 15 or higher.

0	000000000000000000000000000000000000011111111111111111111
0	2222222222222222233333333333333333333333
0	444444444444444444445555555555555555555
0	666666666666666666667777777777777
0	888888888888888889999999999999999
1	000000000000111111111
1	22222222222233333333333
1	444444455
1	66666777
1	8

1.38. **(a)** The first histogram shows two modes: 5–5.2 and 5.6–5.8. **(b)** The second histogram has peaks in locations close to those of the first, but these peaks are much less pronounced, so they would usually be viewed as distinct modes. **(c)** The results will vary with the software used.





1.39. Graph (a) is studying time (Question 4); it is reasonable to expect this to be right-skewed (many students study little or not at all; a few study longer).

Graph (d) is the histogram of student heights (Question 3): One would expect a fair amount of variation but no particular skewness to such a distribution.

The other two graphs are (b) handedness and (c) gender—unless this was a particularly unusual class! We would expect that right-handed students should outnumber lefties substantially. (Roughly 10 to 15% of the population as a whole is left-handed.)

1.40. Sketches will vary. The distribution of coin years would be left-skewed because newer coins are more common than older coins.

1.41. **(a)** Not only are most responses multiples of 10; many are multiples of 30 and 60. Most people will "round" their answers when asked to give an estimate like this; in fact, the most striking answers are ones such as 115, 170, or 230. The students who claimed 360 minutes (6 hours) and 300 minutes (5 hours) may have been exaggerating. (Some students might also "consider suspicious" the student who claimed to study 0 minutes per night. As a teacher, I can easily believe that such students exist, and I suspect that some of your students might easily accept that claim as well.) **(b)** The stemplots suggest that women (claim to) study more than men. The approximate centers are 175 minutes for women and 120 minutes for men.

Women		Men
	0	033334
96	0	66679999
22222221	1	2222222
888888888875555	1	558
4440	2	00344
	2	
	3	0
6	3	

1.42. The stemplot gives more information than a histogram (since all the original numbers can be read off the stemplot), but both give the same impression. The distribution is roughly symmetric with one value (4.88) that is somewhat low. The center of the distribution is between 5.4 and 5.5 (the median is 5.46, the mean is 5.448); if asked to give a single estimate for the "true" density of the earth, something in that range would be the best answer.

48	8
49	
50	7
51	0
52	6799
53	04469
54	2467
55	03578
56	12358
57	59
58	5

1.43. (a) There are four variables: GPA, IQ, and self-concept are quantitative, while gender is categorical. (OBS is not a variable, since it is not really a "characteristic" of a student.) **(b)** Below. **(c)** The distribution is skewed to the left, with center (median) around 7.8. GPAs are spread from 0.5 to 10.8, with only 15 below 6. **(d)** There is more variability among the boys; in fact, there seems to be a subset of boys with GPAs from 0.5 to 4.9. Ignoring that group, the two distributions have similar shapes.

0	5
1	8
2	4
3	4689
4	0679
5	1259
6	0112249
7	22333556666666788899
8	0000222223347899
9	002223344556668
10	01678

Female		Male
	0	5
	1	8
	2	4
4	3	689
7	4	069
952	5	1
4210	6	129
98866533	7	223566666789
997320	8	0002222348
65300	9	2223445668
710	10	68

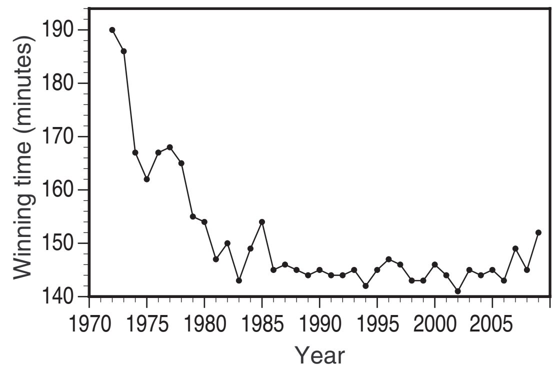
1.44. Stemplot at right, with split stems. The distribution is fairly symmetric—perhaps slightly left-skewed—with center around 110 (clearly above 100). IQs range from the low 70s to the high 130s, with a "gap" in the low 80s.

7	24
7	79
8	
8	69
9	0133
9	6778
10	0022333344
10	555666777789
11	000011112222333444
11	55688999
12	003344
12	677888
13	02
13	6

1.45. Stemplot at right, with split stems. The distribution is skewed to the left, with center around 59.5. Most self-concept scores are between 35 and 73, with a few below that, and one high score of 80 (but not really high enough to be an outlier).

2	01
2	8
3	0
3	5679
4	02344
4	6799
5	1111223344444
5	556668899
6	00001233344444
6	55666677777899
7	0000111223
7	
8	0

1.46. The time plot on the right shows that women's times decreased quite rapidly from 1972 until the mid-1980s. Since that time, they have been fairly consistent: Almost all times since 1986 are between 141 and 147 minutes.



1.47. The total for the 24 countries was 897 days, so with Suriname, it is $897 + 694 = 1591$ days, and the mean is $\bar{x} = \frac{1591}{25} = 63.64$ days.

1.48. The mean score is $\bar{x} = \dfrac{821}{10} = 82.1$.

1.49. To find the ordered list of times, start with the 24 times in Example 1.23, and add 694 to the end of the list. The ordered times (with median highlighted) are

4, 11, 14, 23, 23, 23, 23, 24, 27, 29, 31, 33, 40,
42, 44, 44, 44, 46, 47, 60, 61, 62, 65, 77, 694

The outlier increases the median from 36.5 to 40 days, but the change is much less than the outlier's effect on the mean.

1.50. The median of the service times is 103.5 seconds. (This is the average of the 40th and 41st numbers in the sorted list, but for a set of 80 numbers, we assume that most students will compute the median using software, which does not require that the data be sorted.)

1.51. In order, the scores are:

55, 73, 75, 80, 80, 85, 90, 92, 93, 98

The middle two scores are 80 and 85, so the median is $M = \dfrac{80 + 85}{2} = 82.5$.

1.52. See the ordered list given in the previous solution.
The first quartile is $Q_1 = 75$, the median of the first five numbers: 55, 73, 75, 80, 80. Similarly, $Q_3 = 92$, the median of the last five numbers: 85, 90, 92, 93, 98.

1.53. The maximum and minimum can be found by inspecting the list. The sorted list (with quartile and median locations highlighted) is

1	2	2	3	4	9	9	9	11	19
19	25	30	35	40	44	48	51	52	54
55	56	57	59	64	67	68	73	73	75
75	76	76	77	80	88	89	90	102	103
104	106	115	116	118	121	126	128	137	138
140	141	143	148	148	157	178	179	182	199
201	203	211	225	274	277	289	290	325	367
372	386	438	465	479	700	700	951	1148	2631

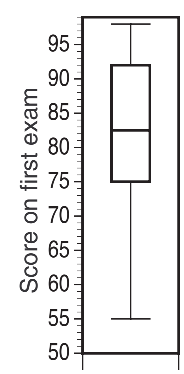
This confirms the five-number summary (1, 54.5, 103.5, 200, and 2631 seconds) given in Example 1.26. The sum of the 80 numbers is 15,726 seconds, so the mean is $\bar{x} = \frac{15{,}726}{80} = 196.575$ seconds (the value 197 in the text was rounded).

Note: *The most tedious part of this process is sorting the numbers and adding them all up. Unless you really want to confirm that your students can sort a list of 80 numbers, consider giving the students the sorted list of times, and checking their ability to identify the locations of the quartiles.*

1.54. The median and quartiles were found earlier; the minimum and maximum are easy to locate in the ordered list of scores (see the solutions to Exercises 1.51 and 1.52), so the five-number summary is Min $= 55$, $Q_1 = 75$, $M = 82.5$, $Q_3 = 92$, Max $= 98$.

1.55. Use the five-number summary from the solution to Exercise 1.54:

$$\text{Min} = 55,\ \ Q_1 = 75,\ \ M = 82.5,\ \ Q_3 = 92,\ \ \text{Max} = 98$$



1.56. The interquartile range is $IQR = Q_3 - Q_1 = 92 - 75 = 17$, so the $1.5 \times IQR$ rule would consider as outliers scores outside the range $Q_1 - 25.5 = 49.5$ to $Q_3 + 25.5 = 117.5$. According to this rule, there are no outliers.

1.57. The variance *can* be computed from the formula $s^2 = \dfrac{1}{n-1}\sum(x_i - \bar{x})^2$; for example, the first term in the sum would be $(80 - 82.1)^2 = 4.41$. However, in practice, software or a calculator is the preferred approach; this yields $s^2 = \dfrac{1416.9}{9} = 157.4\overline{3}$ and $s = \sqrt{s^2} \doteq 12.5472$.

1.58. In order to have $s = 0$, all 5 cases must be equal; for example, 1, 1, 1, 1, 1, or 12.5, 12.5, 12.5, 12.5, 12.5. (If any two numbers are different, then $x_i - \bar{x}$ would be nonzero for some i, so the sum of squared differences would be positive, so $s^2 > 0$, so $s > 0$.)

1.59. Without Suriname, the quartiles are 23 and 46.5 days; with Suriname included, they are 23 and 53.5 days. Therefore, the *IQR* increases from 23.5 to 30.5 days—a much less drastic change than the change in s (18.6 to 132.6 days).

1.60. Divide total score by 4: $\frac{950}{4} = 237.5$ points.

1.61. **(a)** Use a stemplot or histogram. **(b)** Because the distribution is skewed, the five-number summary is the best choice; in millions of dollars, it is

Min	Q_1	M	Q_3	Max
3338	4589	7558.5	13,416	66,667

Some students might choose the less-appropriate summary: $\bar{x} \doteq 12{,}144$ and $s \doteq 12{,}421$ million dollars. **(c)** For example, the distribution is sharply right-skewed. (This is not surprising given that we are looking at the top 100 companies; the top fraction of most distributions will tend to be skewed to the right.)

0	333333333333333333444444444444
0	55555555566666677777777778888889
1	00001112223333333
1	79
2	01111233
2	559
3	114
3	5
4	
4	
5	3
5	99
6	
6	6

1.62. **(a)** Either a stemplot or histogram can be used to display the distribution. Two stemplots are shown on the following page: one with all points, and one with the outlier mentioned in part (b) excluded. In the table are the mean and standard deviation, as well as the five-number summary, both with and without the outlier (all values are percents). The latter is preferable because of the outlier; in particular, note the outlier's effect on the standard deviation. (See also the solution to the next exercise.) **(b)** O'Doul's is marketed as "non-alcoholic" beer.

	$\bar{x}$	s	Min	Q_1	M	Q_3	Max
All points	4.7593	0.7523	0.4	4.30	4.7	5	6.5
No O'Doul's	4.8106	0.5864	3.8	4.35	4.7	5	6.5

Note: *In federal regulations, part of the definition of beer is that it has at least 0.5% alcohol. By that standard, O'Doul's is a low-alcohol* beverage, *but it is not* beer.

All points

0	4
0	
1	
1	
2	
2	
3	
3	88
4	1111112222222222223334444
4	55555566666777777777888999999999999
5	000000011224
5	5666688999999
6	1
6	5

Without O'Doul's

3	88
4	111111
4	2222222222333
4	4444555555
4	66666777777777
4	8889999999999
5	000000011
5	22
5	45
5	6666
5	88999999
6	1
6	
6	5

1.63. All of these numbers are given in the table in the solution to the previous exercise. **(a)** $\bar{x}$ changes from 4.76% (with) to 4.81% (without); the median (4.7%) does not change. **(b)** s changes from 0.7523% to 0.5864%; Q_1 changes from 4.3% to 4.35%, while $Q_3 = 5\%$ does not change. **(c)** A low outlier decreases $\bar{x}$; any kind of outlier increases s. Outliers have little or no effect on the median and quartiles.

1.64. **(a)** A stemplot or histogram can be used to display the distribution. Students may report either mean/standard deviation or the five-number summary (in units of calories):

$\bar{x}$	s	Min	Q_1	M	Q_3	Max
141.06	27.79	70	113	145.5	157	210

(b) O'Doul's has the fewest calories (70) of these 86 beers. **(c)** Nearly all the beers with fewer than 120 calories are marketed as light beers (and most have "light" in their names). Of the other beers, only one (Weinhard's Amber Light) is called "light."

7	0
8	
9	4556889
10	2458
11	00000000334
12	08
13	0235558
14	22333444555666788899
15	0012233356777
16	00012336669
17	01459
18	8
19	5
20	00
21	0

Note: *If we apply the $1.5 \times IQR$ rule to all 86 beers, O'Doul's does not qualify as an outlier (the cutoff is 47). However, if we restrict our attention to the light beers (fewer than 120 calories), any beer below 80 calories is an outlier.*

1.65. Use a small data set with an odd number of points, so that the median is the middle number. After deleting the lowest observation, the median will be the average of that middle number and the next number after it; if that latter number is much larger, the median will change substantially. For example, start with 0, 1, $\boxed{2}$, 998, 1000; after removing 0, the median changes from 2 to 500.

1.66. Salary distributions (especially in professional sports) tend to be skewed to the right. This skew makes the mean higher than the median.

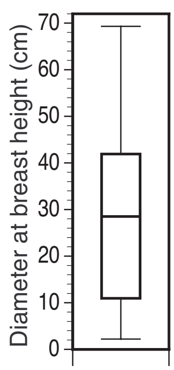
1.67. **(a)** The distribution is left-skewed. While the skew makes the five-number summary is preferable, some students might give the mean/standard deviation. In ounces, these statistics are:

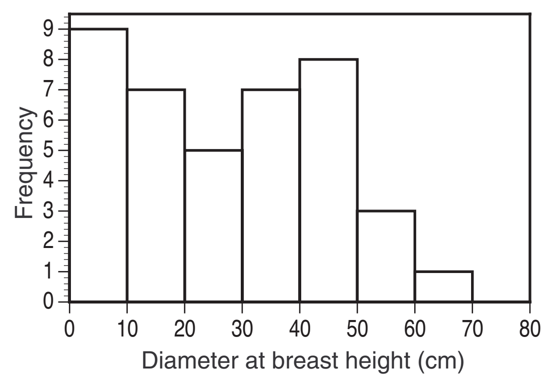
$\bar{x}$	s	Min	Q_1	M	Q_3	Max
6.456	1.425	3.7	4.95	6.7	7.85	8.2

3	7
4	3
4	7777
5	23
5	
6	0033
6	7
7	03
7	668899999
8	2

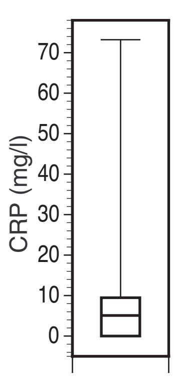
(b) The numerical summary does not reveal the two weight clusters (visible in a stemplot or histogram). **(c)** For small potatoes (less than 6 oz), $n = 8$, $\bar{x} = 4.662$ oz, and $s = 0.501$ oz. For large potatoes, $n = 17$, $\bar{x} = 7.300$ oz, and $s = 0.755$ oz. Because there are clearly two groups, it seems appropriate to treat them separately.

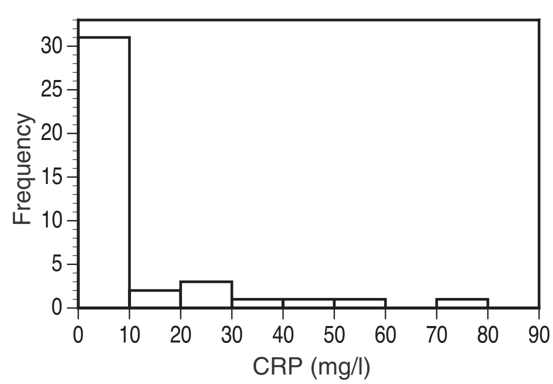
1.68. **(a)** The five-number summary is Min = 2.2 cm, $Q_1 = 10.95$ cm, $M = 28.5$ cm, $Q_3 = 41.9$ cm, Max = 69.3 cm. **(b)** & **(c)** The boxplot and histogram are shown below. (Students might choose different interval widths for the histogram.) **(d)** Preferences will vary. Both plots reveal the right-skew of this distribution, but the boxplot does not show the two peaks visible in the histogram.





1.69. **(a)** The five-number summary is Min = 0 mg/l, $Q_1 = 0$ mg/l, $M = 5.085$ mg/l, $Q_3 = 9.47$ mg/l, Max = 73.2 mg/l. **(b)** & **(c)** The boxplot and histogram are shown below. (Students might choose different interval widths for the histogram.) **(d)** Preferences will vary. Both plots reveal the sharp right-skew of this distribution, but because Min = Q_1, the boxplot looks somewhat strange. The histogram seems to convey the distribution better.

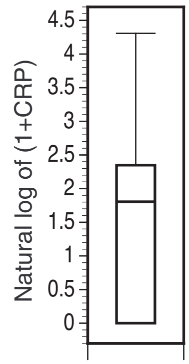


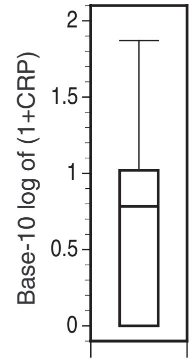


1.70. Answers depend on whether natural (base-e) or common (base-10) logarithms are used. Both sets of answers are shown here. If this exercise is assigned, it would probably be best for the sanity of both instructor and students to specify which logarithm to use.

(a) The five-number summary is:

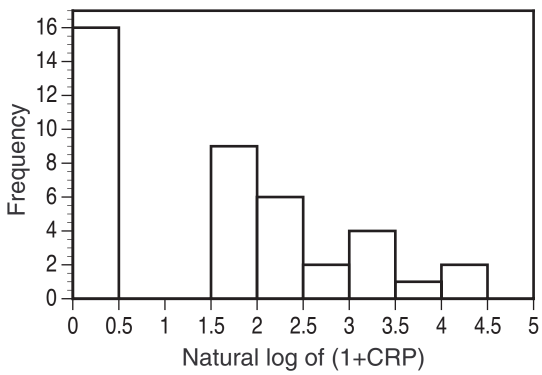
Logarithm	Min	Q_1	M	Q_3	Max
Natural	0	0	1.8048	2.3485	4.3068
Common	0	0	0.7838	1.0199	1.8704

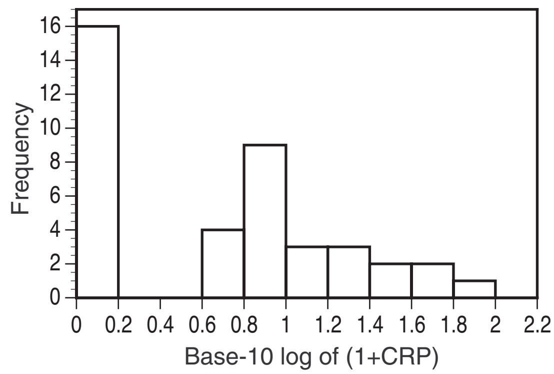




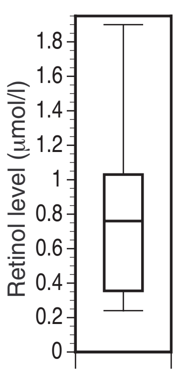
(The ratio between these answers is roughly $\ln 10 \doteq 2.3$.)

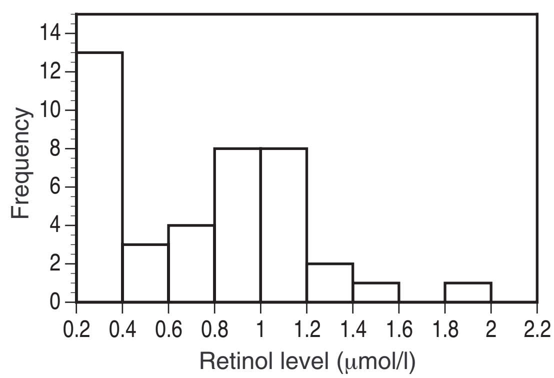
(b) & **(c)** The boxplots and histograms are shown below. (Students might choose different interval widths for the histograms.) **(d)** As for Exercise 1.69, preferences will vary.





1.71. **(a)** The five-number summary (in units of μmol/l) is Min $= 0.24$, $Q_1 = 0.355$, $M = 0.76$, $Q_3 = 1.03$, Max $= 1.9$. **(b)** & **(c)** The boxplot and histogram are shown below. (Students might choose different interval widths for the histogram.) **(d)** The distribution is right-skewed. A histogram (or stemplot) is preferable because it reveals an important feature not evident from a boxplot: This distribution has two peaks.





1.72. The mean and standard deviation for these ratings are $\bar{x} = 5.9$ and $s \doteq 3.7719$; the five-number summary is Min $= Q_1 = 1$, $M = 6.5$, $Q_3 =$ Max $= 10$. For a graphical presentation, a stemplot (or histogram) is better than a boxplot because the latter obscures details about the distribution. (With a little thought, one might realize that Min $= Q_1 = 1$ and $Q_3 =$ Max $= 10$ means that there are lots of 1's and lots of 10's, but this is much more evident in a stemplot or histogram.)

1	0000000000000000
2	0000
3	0
4	0
5	00000
6	000
7	0
8	000000
9	00000
10	000000000000000000

1.73. The distribution of household net worth would almost surely be strongly skewed to the right: Most families would generally have accumulated little or modest wealth, but a few would have become rich. This strong skew pulls the mean to be higher than the median.

1.74. See also the solution to Exercise 1.36. **(a)** The five-number summary (in units of metric tons per person) is:

Min $= 0$, $Q_1 = 0.75$, $M = 3.2$, $Q_3 = 7.8$, Max $= 19.9$

The evidence for the skew is in the large gaps between the higher numbers; that is, the differences $Q_3 - M$ and Max $- Q_3$ are large compared to $Q_1 -$ Min and $M - Q_1$. **(b)** The *IQR* is $Q_3 - Q_1 = 7.05$, so outliers would be less than -9.825 or greater than 18.375. According to this rule, only the United States qualifies as an outlier, but Canada and Australia seem high enough to also include them.

0	00000000000000011111
0	222233333
0	445
0	6677
0	888999
1	001
1	
1	
1	67
1	9

1.75. The total salary is \$690,000, so the mean is $\bar{x} = \frac{\$690{,}000}{9} \doteq \$76{,}667$. Six of the nine employees earn less than the mean. The median is $M = \$35{,}000$.

1.76. If three individuals earn \$0, \$0, and \$20,000, the reported median is \$20,000. If the two individuals with no income take jobs at \$14,000 each, the median decreases to \$14,000. The same thing can happen to the mean: In this example, the mean drops from \$20,000 to \$16,000.

1.77. The total salary is now \$825,000, so the new mean is $\bar{x} = \frac{\$825{,}000}{9} \doteq \$91{,}667$. The median is unchanged.

1.78. Details at right.

$$\bar{x} = \frac{11{,}200}{7} = 1600$$

$$s^2 = \frac{214{,}872}{6} = 35{,}812 \text{ and}$$

$$s = \sqrt{35{,}812} \doteq 189.24$$

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
1792	192	36864
1666	66	4356
1362	−238	56644
1614	14	196
1460	−140	19600
1867	267	71289
1439	−161	25921
11200	0	214872

1.79. The quote describes a distribution with a strong right skew: Lots of years with no losses to hurricane (\$0), but very high numbers when they do occur. For example, if there is one hurricane in a 10-year period causing \$1 million in damages, the "average annual loss" for that period would be \$100,000, but that does not adequately represent the cost for the year of the hurricane. Means are not the appropriate measure of center for skewed distributions.

1.80. (a) $\bar{x}$ and s are appropriate for symmetric distributions with no outliers. **(b)** Both high numbers are flagged as outliers. For women, $IQR = 60$, so the upper $1.5 \times IQR$ limit is 300 minutes. For men, $IQR = 90$, so the upper $1.5 \times IQR$ limit is 285 minutes. The table on the right shows the effect of removing these outliers.

	Women		Men	
	$\bar{x}$	s	$\bar{x}$	s
Before	165.2	56.5	117.2	74.2
After	158.4	43.7	110.9	66.9

1.81. (a) & **(b)** See the table on the right. In both cases, the mean and median are quite similar.

	$\bar{x}$	s	M
pH	5.4256	0.5379	5.44
Density	5.4479	0.2209	5.46

1.82. See also the solution to Exercise 1.43. **(a)** The mean of this distribution appears to be higher than 100. (There is no substantial difference between the standard deviations.) **(b)** The mean and median are quite similar; the mean is slightly smaller due to the slight left skew of the data. **(c)** In addition to the mean and median, the standard deviation is shown for reference (the exercise did not ask for it).

	$\bar{x}$	s	M
IQ	108.9	13.17	110
GPA	7.447	(2.1)	7.829

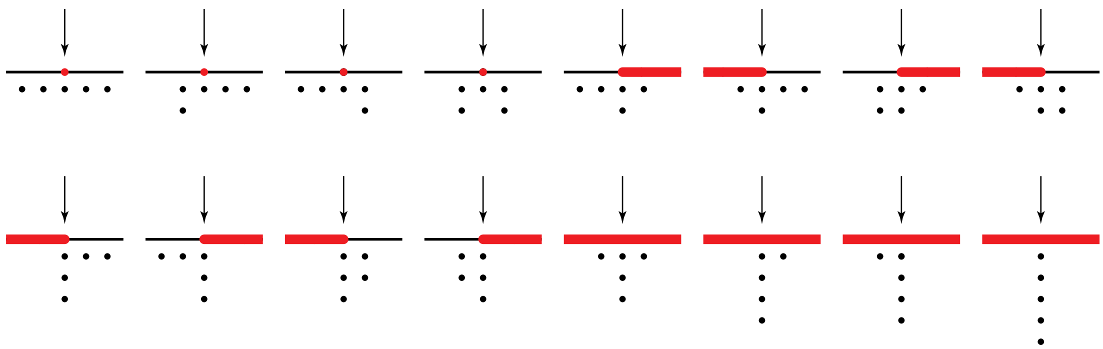
Note: *Students may be somewhat puzzled by the statement in (b) that the median is "close to the mean" (when they differ by 1.1), followed by (c), where they "differ a bit" (when $M - \bar{x} = 0.382$). It may be useful to emphasize that we judge the size of such differences relative to the spread of the distribution. For example, we can note that $\frac{1.1}{13.17} \doteq 0.08$ for (b), and $\frac{0.382}{2.1} \doteq 0.18$ for (c).*

1.83. With only two observations, the mean and median are always equal because the median is halfway between the middle two (in this case, the only two) numbers.

1.84. (a) The mean (green arrow) moves along with the moving point (in fact, it moves in the same direction as the moving point, at one-third the speed). At the same time, as long as the moving point remains to the right of the other two, the median (red arrow) points to the middle point (the rightmost nonmoving point). **(b)** The mean follows the moving point as before. When the moving point passes the rightmost fixed point, the median slides along with it until the moving point passes the leftmost fixed point, then the median stays there.

1.85. (a) There are several different answers, depending on the configuration of the first five points. *Most students* will likely assume that the first five points should be distinct (no repeats), in which case the sixth point *must* be placed at the median. This is because the median of 5 (sorted) points is the third, while the median of 6 points is the average of the third and fourth. If these are to be the same, the third and fourth points of the set of six must both equal the third point of the set of five.

The diagram below illustrates all of the possibilities; in each case, the arrow shows the location of the median of the initial five points, and the shaded region (or dot) on the line indicates where the sixth point can be placed without changing the median. Notice that there are four cases where the median does not change, regardless of the location of the sixth point. (The points need not be equally spaced; these diagrams were drawn that way for convenience.)

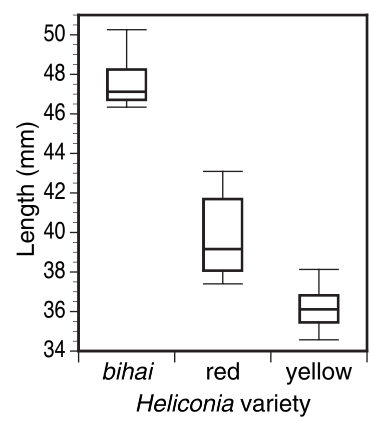


(b) Regardless of the configuration of the first five points, if the sixth point is added so as to leave the median unchanged, then in that (sorted) set of six, the third and fourth points must be equal. One of these two points will be the middle (fourth) point of the (sorted) set of seven, no matter where the seventh point is placed.

Note: *If you have a student who illustrates all possible cases above, then it is likely that the student either (1) obtained a copy of this solutions manual, (2) should consider a career in writing solutions manuals, (3) has too much time on his or her hands, or (4) both 2 and 3 (and perhaps 1) are true.*

1.86. The five-number summaries (all in millimeters) are:

	Min	Q_1	M	Q_3	Max
bihai	46.34	46.71	47.12	48.245	50.26
red	37.40	38.07	39.16	41.69	43.09
yellow	34.57	35.45	36.11	36.82	38.13



H. bihai is clearly the tallest variety—the shortest *bihai* was over 3 mm taller than the tallest red. Red is generally taller than yellow, with a few exceptions. Another noteworthy fact: The red variety is more variable than either of the other varieties.

1.87. (a) The means and standard deviations (all in millimeters) are:

Variety	$\bar{x}$	s
bihai	47.5975	1.2129
red	39.7113	1.7988
yellow	36.1800	0.9753

bihai

46	3466789
47	114
48	0133
49	
50	12

red

37	4789
38	0012278
39	167
40	56
41	4699
42	01
43	0

yellow

34	56
35	146
36	0015678
37	01
38	1

(b) *Bihai* and red appear to be right-skewed (although it is difficult to tell with such small samples). Skewness would make these distributions unsuitable for $\bar{x}$ and s.

1.88. **(a)** The mean is $\bar{x} = 15$, and the standard deviation is $s \doteq 5.4365$. **(b)** The mean is still 15; the new standard deviation is 3.7417. **(c)** Using the mean as a substitute for missing data will not change the mean, but it decreases the standard deviation.

1.89. The minimum and maximum are easily determined to be 1 and 12 letters, and the quartiles and median can be found by adding up the bar heights. For example, the first two bars have total height 22.3% (less than 25%), and adding the third bar brings the total to 45%, so Q_1 must equal 3 letters. Continuing this way, we find that the five-number summary, in units of letters, is:

$$\text{Min} = 1,\ \ Q_1 = 3,\ \ M = 4,\ \ Q_3 = 5,\ \ \text{Max} = 12$$

Note that even without the frequency table given in the data file, we could draw the same conclusion by estimating the heights of the bars in the histogram.

1.90. Because the mean is to be 7, the five numbers must add up to 35. Also, the third number (in order from smallest to largest) must be 10 because that is the median. Beyond that, there is some freedom in how the numbers are chosen.

Note: *It is likely that many students will interpret "positive numbers" as meaning positive integers only, which leads to eight possible solutions, shown below.*

1	1	10	10	13	1	1	10	11	12	1	2	10	10	12	1	2	10	11	11
1	3	10	10	11	1	4	10	10	10	2	2	10	10	11	2	3	10	10	10

1.91. The simplest approach is to take (at least) six numbers—say, a, b, c, d, e, f in increasing order. For this set, $Q_3 = e$; we can cause the mean to be larger than e by simply choosing f to be *much* larger than e. For example, if all numbers are nonnegative, $f > 5e$ would accomplish the goal because then

$$\bar{x} = \frac{a+b+c+d+e+f}{6} > \frac{e+f}{6} > \frac{e+5e}{6} = e.$$

1.92. The algebra might be a bit of a stretch for some students:

$$\begin{aligned}
& (x_1 - \bar{x}) + (x_2 - \bar{x}) + (x_3 - \bar{x}) + \cdots + (x_{n-1} - \bar{x}) + (x_n - \bar{x}) \\
= {} & x_1 - \bar{x} + x_2 - \bar{x} + x_3 - \bar{x} + \cdots + x_{n-1} - \bar{x} + x_n - \bar{x} \\
& \qquad \textit{(drop all the parentheses)} \\
= {} & x_1 + x_2 + x_3 + \cdots + x_{n-1} + x_n - \bar{x} - \bar{x} - \bar{x} - \cdots - \bar{x} - \bar{x} \\
& \qquad \textit{(rearrange the terms)} \\
= {} & x_1 + x_2 + x_3 + \cdots + x_{n-1} + x_n - n \cdot \bar{x}
\end{aligned}$$

Next, simply observe that $n \cdot \bar{x} = x_1 + x_2 + x_3 + \cdots + x_{n-1} + x_n$.

1.93. **(a)** One possible answer is 1, 1, 1, 1. **(b)** 0, 0, 20, 20. **(c)** For (a), any set of four identical numbers will have $s = 0$. For (b), the answer is unique; here is a rough description of why. We want to maximize the "spread-out"-ness of the numbers (which is what standard deviation measures), so 0 and 20 seem to be reasonable choices based on that idea. We also want to make each individual squared deviation—$(x_1 - \bar{x})^2$, $(x_2 - \bar{x})^2$, $(x_3 - \bar{x})^2$, and $(x_4 - \bar{x})^2$—as large as possible. If we choose 0, 20, 20, 20—or 20, 0, 0, 0—we make the first squared deviation 15^2, but the other three are only 5^2. Our best choice is two at each extreme, which makes all four squared deviations equal to 10^2.

1.94. Answers will vary. Typical calculators will carry only about 12 to 15 digits; for example, a TI-83 fails (gives $s = 0$) for 14-digit numbers. *Excel* (at least the version I checked) also fails for 14-digit numbers, but it gives $s = 262{,}144$ rather than 0. The (very old) version of Minitab used to prepare these answers fails at 20,000,001 (eight digits), giving $s = 2$.

1.95. The table on the right reproduces the means and standard deviations from the solution to Exercise 1.87 and shows those values expressed in inches. For each conversion, multiply by $39.37/1000 = 0.03937$ (or divide by 25.4—an inch is defined as 25.4 millimeters). For example, for the *bihai* variety, $\bar{x} = (47.5975 \text{ mm})(0.03937 \text{ in/mm}) = (47.5975 \text{ mm}) \div (25.4 \text{ mm/in}) = 1.874$ in.

	(in mm)		(in inches)	
Variety	$\bar{x}$	s	$\bar{x}$	s
bihai	47.5975	1.2129	1.874	0.04775
red	39.7113	1.7988	1.563	0.07082
yellow	36.1800	0.9753	1.424	0.03840

1.96. **(a)** $\bar{x} = 5.4479$ and $s = 0.2209$. **(b)** The first measurement corresponds to $5.50 \times 62.43 = 343.365$ pounds per cubic foot. To find $\bar{x}_{\text{new}}$ and s_{new}, we similarly multiply by 62.43: $\bar{x}_{\text{new}} \doteq 340.11$ and $s_{\text{new}} \doteq 13.79$.

Note: *The conversion from cm to feet is included in the multiplication by 62.43; the step-by-step process of this conversion looks like this:*

$$(1 \text{ g/cm}^3)(0.001 \text{ kg/g})(2.2046 \text{ lb/kg})(30.48^3 \text{ cm}^3/\text{ft}^3) = 62.43 \text{ lb/ft}^3$$

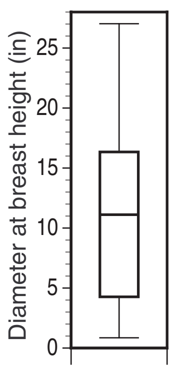
1.97. Convert from kilograms to pounds by multiplying by 2.2: $\bar{x} = (2.42 \text{ kg})(2.2 \text{ lb/kg}) \doteq 5.32$ lb and $s = (1.18 \text{ kg})(2.2 \text{ lb/kg}) \doteq 2.60$ lb.

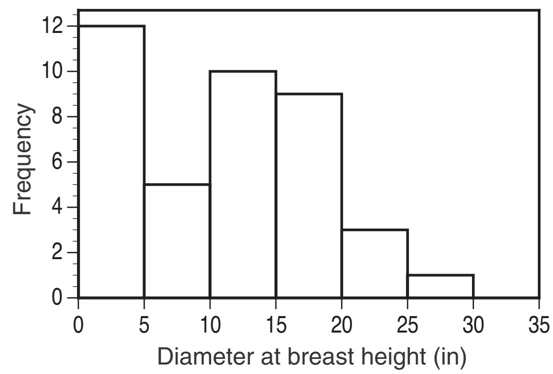
1.98. Variance is changed by a factor of $2.54^2 = 6.4516$; generally, for a transformation $x_{\text{new}} = a + bx$, the new variance is b^2 times the old variance.

1.99. There are 80 service times, so to find the 10% trimmed mean, remove the highest and lowest eight values (leaving 64). Remove the highest and lowest 16 values (leaving 48) for the 20% trimmed mean.

The mean and median for the full data set are $\bar{x} = 196.575$ and $M = 103.5$ minutes. The 10% trimmed mean is $\bar{x}^* \doteq 127.734$, and the 20% trimmed mean is $\bar{x}^{**} \doteq 111.917$ minutes. Because the distribution is right-skewed, removing the extremes lowers the mean.

1.100. After changing the scale from centimeters to inches, the five-number summary values change by the same ratio (that is, they are multiplied by 0.39). The shape of the histogram might change slightly because of the change in class intervals. **(a)** The five-number summary (in inches) is Min $= 0.858$, $Q_1 = 4.2705$, $M = 11.115$, $Q_3 = 16.341$, Max $= 27.027$. **(b)** & **(c)** The boxplot and histogram are shown below. (Students might choose different interval widths for the histogram.) **(d)** As in Exercise 1.56, the histogram reveals more detail about the shape of the distribution.





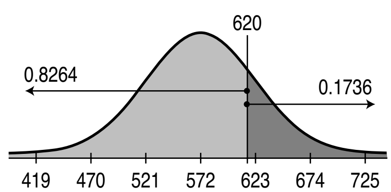
1.101. Take the mean plus or minus two standard deviations: $572 \pm 2(51) = 470$ to 674.

1.102. Take the mean plus or minus three standard deviations: $572 \pm 3(51) = 419$ to 725.

1.103. The z-score is $z = \frac{620-572}{51} \doteq 0.94$.

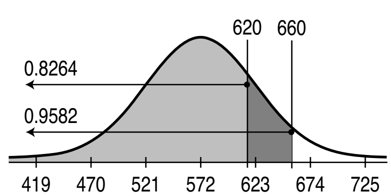
1.104. The z-score is $z = \frac{510-572}{51} \doteq -1.22$. This is negative because an ISTEP score of 510 is below average; specifically, it is 1.22 standard deviations below the mean.

1.105. Using Table A, the proportion below 620 ($z \doteq 0.94$) is 0.8264 and the proportion at or above is 0.1736; these two proportions add to 1. The graph on the right illustrates this with a single curve; it conveys essentially the same idea as the "graphical subtraction" picture shown in Example 1.36.



1.106. Using Table A, the proportion below 620 ($z \doteq 0.94$) is 0.8264, and the proportion below 660 ($z \doteq 1.73$) is 0.9582. Therefore:

$$\begin{array}{ccccc} \text{area between} & & \text{area left} & & \text{area left} \\ \text{620 and 660} & = & \text{of 660} & - & \text{of 620} \\ \\ 0.1318 & = & 0.9582 & - & 0.8264 \end{array}$$

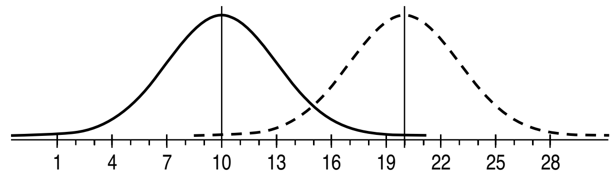


The graph on the right illustrates this with a single curve; it conveys essentially the same idea as the "graphical subtraction" picture shown in Example 1.37.

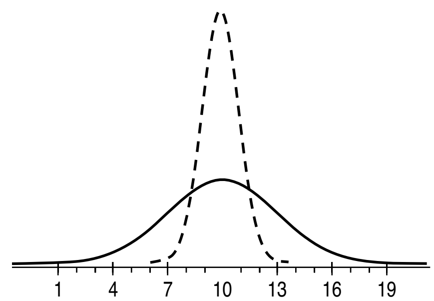
1.107. Using Table A, this ISTEP score should correspond to a standard score of $z \doteq 0.67$ (software gives 0.6745), so the ISTEP score (unstandardized) is $572 + 0.67(51) \doteq 606.2$ (software: 606.4).

1.108. Using Table A, x should correspond to a standard score of $z \doteq -0.84$ (software gives -0.8416), so the ISTEP score (unstandardized) is $x = 572 - 0.84(51) \doteq 529.2$ (software: 529.1).

1.109. Of course, student sketches will not be as neat as the curves on the right, but they should have roughly the correct shape. **(a)** It is easiest to draw the curve first, and then mark the scale on the axis. **(b)** Draw a copy of the first curve, with the peak over 20. **(c)** The curve has the same shape, but is translated left or right.



1.110. **(a)** As in the previous exercise, draw the curve first, and then mark the scale on the axis. **(b)** In order to have a standard deviation of 1, the curve should be 1/3 as wide, and three times taller. **(c)** The curve is centered at the same place (the mean), but its height and width change. Specifically, increasing the standard deviation makes the curve wider and shorter; decreasing the standard deviation makes the curve narrower and taller.



1.111. Sketches will vary.

1.112. **(a)** The table on the right gives the ranges for women; for example, about 68% of women speak between 7856 and 20,738 words per day. **(b)** Negative numbers do not make sense for this situation. The 68–95–99.7 rule is reasonable for a distribution that is close to Normal, but by constructing a stemplot or histogram, it is easily confirmed that this distribution is slightly right-skewed. **(c)** These ranges are also in the table; the men's distribution is more skewed than the women's distribution, so the 68–95–99.7 rule is even less appropriate. **(d)** This does not support the conventional wisdom: The ranges from parts (a) and (c) overlap quite a bit. Additionally, the difference in the means is quite small relative to the large standard deviations.

	Women	Men
68%	7856 to 20,738	4995 to 23,125
95%	1415 to 27,179	−4070 to 32,190
99.7%	−5026 to 33,620	−13,135 to 41,255

1.113. **(a)** Ranges are given in the table on the right. In both cases, some of the lower limits are negative, which does not make sense; this happens because the women's distribution is skewed, and the men's distribution has an outlier. Contrary to the conventional wisdom, the men's mean is slightly higher, although the outlier is at least partly responsible for that. **(b)** The means suggest that Mexican men and women tend to speak more than people of the same gender from the United States.

	Women	Men
68%	8489 to 20,919	7158 to 22,886
95%	2274 to 27,134	−706 to 30,750
99.7%	−3941 to 33,349	−8,570 to 38,614

1.114. **(a)** For example, $\frac{68-70}{10} = -0.2$. The complete list is given on the right. **(b)** The cut-off for an A is the 85th percentile for the $N(0, 1)$ distribution. From Table A, this is approximately 1.04; software gives 1.0364. **(c)** The top two students (with scores of 92 and 98) received A's.

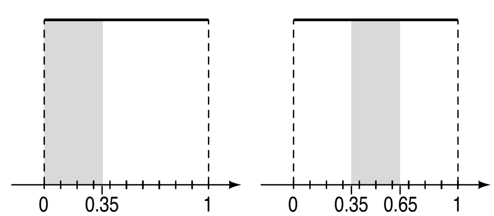
68	−0.2
54	−1.6
92	2.2
75	0.5
73	0.3
98	2.8
64	−0.6
55	−1.5
80	1
70	0

1.115. **(a)** We need the 5th, 15th, 55th, and 85th percentiles for a $N(0, 1)$ distribution. These are given in the table on the right. **(b)** To convert to actual scores, take the standard-score cut-off z and compute $10z + 70$. **(c)** Opinions will vary.

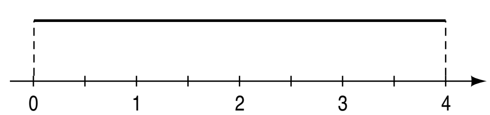
	Table A		Software	
	Standard	Actual	Standard	Actual
F	−1.64	53.6	−1.6449	53.55
D	−1.04	59.6	−1.0364	59.64
C	0.13	71.3	0.1257	71.26
B	1.04	80.4	1.0364	80.36

Note: *The cut-off for an A given in the previous solution is the* lowest *score that gets an A—that is, the point where one's grade drops from an A to a B. These cut-offs are the points where one's grade jumps* up. *In practice, this is only an issue for a score that falls exactly on the border between two grades.*

1.116. **(a)** The curve forms a 1×1 square, which has area 1.
(b) $P(X < 0.35) = 0.35$.
(c) $P(0.35 < X < 0.65) = 0.3$.



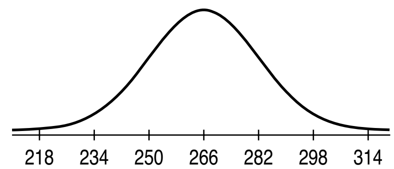
1.117. **(a)** The height should be $\frac{1}{4}$ since the area under the curve must be 1. The density curve is on the right. **(b)** $P(X \leq 1) = \frac{1}{4} = 0.25$. **(c)** $P(0.5 < X < 2.5) = 0.5$.



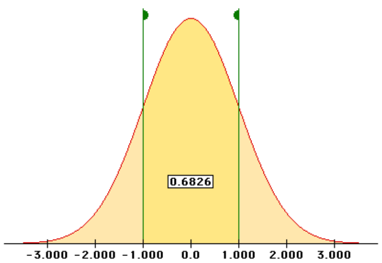
1.118. The mean and median both equal 0.5; the quartiles are $Q_1 = 0.25$ and $Q_3 = 0.75$.

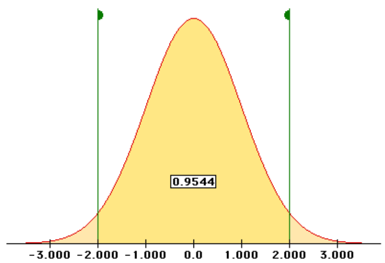
1.119. **(a)** Mean is C, median is B (the right skew pulls the mean to the right). **(b)** Mean A, median A. **(c)** Mean A, median B (the left skew pulls the mean to the left).

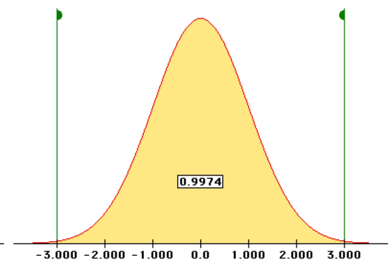
1.120. Hint: It is best to draw the curve first, then place the numbers below it. Students may at first make mistakes like drawing a half-circle instead of the correct "bell-shaped" curve, or being careless about locating the standard deviation.



1.121. **(a)** The applet shows an area of 0.6826 between −1.000 and 1.000, while the 68–95–99.7 rule rounds this to 0.68. **(b)** Between −2.000 and 2.000, the applet reports 0.9544 (compared to the rounded 0.95 from the 68–95–99.7 rule). Between −3.000 and 3.000, the applet reports 0.9974 (compared to the rounded 0.997).

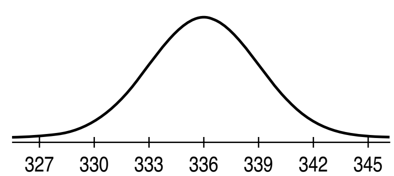






1.122. See the sketch of the curve in the solution to Exercise 1.120. **(a)** The middle 95% fall within two standard deviations of the mean: $266 \pm 2(16)$, or 234 to 298 days. **(b)** The shortest 2.5% of pregnancies are shorter than 234 days (more than two standard deviations below the mean).

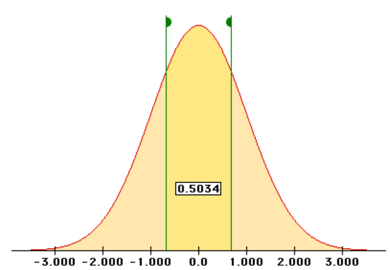
1.123. **(a)** 99.7% of horse pregnancies fall within three standard deviations of the mean: $336 \pm 3(3)$, or 327 to 325 days. **(b)** About 16% are longer than 339 days since 339 days or more corresponds to at least one standard deviation above the mean.



Note: *This exercise did not ask for a sketch of the Normal curve, but students should be encouraged to make such sketches anyway.*

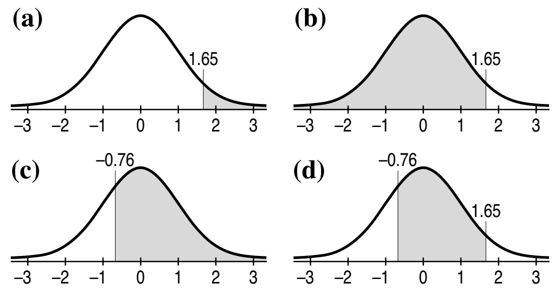
1.124. Because the quartiles of any distribution have 50% of observations between them, we seek to place the flags so that the reported area is 0.5. The closest the applet gets is an area of 0.5034, between −0.680 and 0.680. Thus, the quartiles of any Normal distribution are about 0.68 standard deviations above and below the mean.

Note: *Table A places the quartiles at about ±0.67; other statistical software gives ±0.6745.*

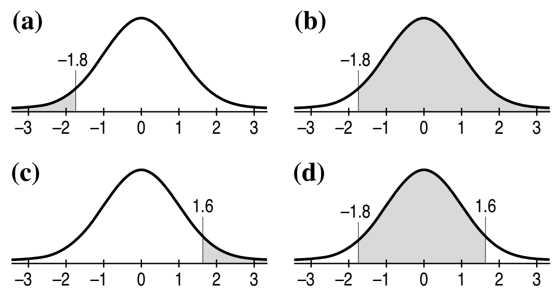


1.125. The mean and standard deviation are $\bar{x} = 5.4256$ and $s = 0.5379$. About 67.62% ($71/105 \doteq 0.6476$) of the pH measurements are in the range $\bar{x} \pm s = 4.89$ to 5.96. About 95.24% (100/105) are in the range $\bar{x} \pm 2s = 4.35$ to 6.50. All (100%) are in the range $\bar{x} \pm 3s = 3.81$ to 7.04.

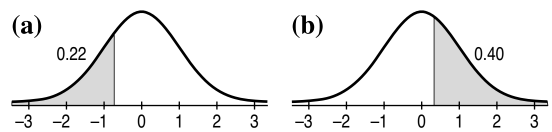
1.126. Using values from Table A:
(a) $Z > 1.65$: 0.0495. **(b)** $Z < 1.65$: 0.9505. **(c)** $Z > -0.76$: 0.7764. **(d)** $-0.76 < Z < 1.65$: $0.9505 - 0.2236 = 0.7269$.



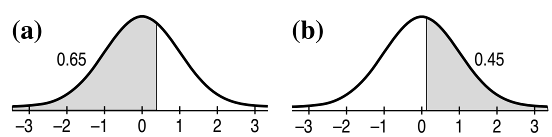
1.127. Using values from Table A:
(a) $Z \leq -1.8$: 0.0359. **(b)** $Z \geq -1.8$: 0.9641. **(c)** $Z > 1.6$: 0.0548. **(d)** $-1.8 < Z < 1.6$: $0.9452 - 0.0359 = 0.9093$.



1.128. **(a)** 22% of the observations fall below -0.7722. (This is the 22nd percentile of the standard Normal distribution.) **(b)** 40% of the observations fall above 0.2533 (the 60th percentile of the standard Normal distribution).



1.129. **(a)** $z = 0.3853$ has cumulative proportion 0.65 (that is, 0.3853 is the 65th percentile of the standard Normal distribution). **(b)** If $z = 0.1257$, then $Z > z$ has proportion 0.45 (0.1257 is the 55th percentile).



1.130. 70 is two standard deviations below the mean (that is, it has standard score $z = -2$), so about 2.5% (half of the outer 5%) of adults would have WAIS scores below 70.

1.131. 130 is two standard deviations above the mean (that is, it has standard score $z = 2$), so about 2.5% of adults would score at least 130.

1.132. Tonya's score standardizes to $z = \frac{1820-1509}{321} \doteq 0.9688$, while Jermaine's score corresponds to $z = \frac{29-21.5}{5.4} \doteq 1.3889$. Jermaine's score is higher.

1.133. Jacob's score standardizes to $z = \frac{16-21.5}{5.4} \doteq -1.0185$, while Emily's score corresponds to $z = \frac{1020-1509}{321} \doteq -1.5234$. Jacob's score is higher.

1.134. Jose's score standardizes to $z = \frac{2080-1509}{321} \doteq 1.7788$, so an equivalent ACT score is $21.5 + 1.7788 \times 5.4 \doteq 31.1$. (Of course, ACT scores are reported as whole numbers, so this would presumably be a score of 31.)

1.135. Maria's score standardizes to $z = \frac{30-21.5}{5.4} \doteq 1.5741$, so an equivalent SAT score is $1509 + 1.5741 \times 321 \doteq 2014$.

1.136. Maria's score standardizes to $z = \frac{2090-1509}{321} \doteq 1.81$, for which Table A gives 0.9649. Her score is the 96.5 percentile.

1.137. Jacob's score standardizes to $z = \frac{19-21.5}{5.4} \doteq -0.4630$, for which Table A gives 0.3228. His score is the 32.3 percentile.

1.138. 1920 and above: The top 10% corresponds to a standard score of $z = 1.2816$, which in turn corresponds to a score of $1509 + 1.2816 \times 321 \doteq 1920$ on the SAT.

1.139. 1239 and below: The bottom 20% corresponds to a standard score of $z = -0.8416$, which in turn corresponds to a score of $1509 - 0.8416 \times 321 \doteq 1239$ on the SAT.

1.140. The quartiles of a Normal distribution are ± 0.6745 standard deviations from the mean, so for ACT scores, they are $21.5 \pm 0.6745 \times 5.4 \doteq 17.9$ to 25.1.

1.141. The quintiles of the SAT score distribution are $1509 - 0.8416 \times 321 = 1239$, $1509 - 0.2533 \times 321 = 1428$, $1509 + 0.2533 \times 321 = 1590$, and $1509 + 0.8416 \times 321 = 1779$.

1.142. For a Normal distribution with mean 55 mg/dl and standard deviation 15.5 mg/dl: **(a)** 40 mg/dl standardizes to $z = \frac{40-55}{15.5} \doteq -0.9677$. Using Table A, 16.60% of women fall below this level (software: 16.66%). **(b)** 60 mg/dl standardizes to $z = \frac{60-55}{15.5} \doteq 0.3226$. Using Table A, 37.45**(c)** Subtract the answers from (a) and (b) from 100%: Table A gives 45.95% (software: 45.99%), so about 46% of women fall in the intermediate range.

1.143. For a Normal distribution with mean 46 mg/dl and standard deviation 13.6 mg/dl: **(a)** 40 mg/dl standardizes to $z = \frac{40-46}{13.6} \doteq -0.4412$. Using Table A, 33% of men fall below this level (software: 32.95%). **(b)** 60 mg/dl standardizes to $z = \frac{60-46}{13.6} \doteq 1.0294$. Using Table A, 15.15**(c)** Subtract the answers from (a) and (b) from 100%: Table A gives 51.85% (software: 51.88%), so about 52% of men fall in the intermediate range.

1.144. **(a)** About 0.6% of healthy young adults have osteoporosis (the cumulative probability below a standard score of -2.5 is 0.0062). **(b)** About 31% of this population of older women has osteoporosis: The BMD level which is 2.5 standard deviations below the young adult mean would standardize to -0.5 for these older women, and the cumulative probability for this standard score is 0.3085.

1.145. **(a)** About 5.2%: $x < 240$ corresponds to $z < -1.625$. Table A gives 5.16% for -1.63 and 5.26% for -1.62. Software (or averaging the two table values) gives 5.21%. **(b)** About 54.7%: $240 < x < 270$ corresponds to $-1.625 < z < 0.25$. The area to the left of 0.25 is 0.5987; subtracting the answer from part (a) leaves about 54.7%. **(c)** About 279 days or longer: Searching Table A for 0.80 leads to $z > 0.84$, which corresponds to $x > 266 + 0.84(16) = 279.44$. (Using the software value $z > 0.8416$ gives $x > 279.47$.)

1.146. **(a)** The quartiles for a standard Normal distribution are ± 0.6745. **(b)** For a $N(\mu, \sigma)$ distribution, $Q_1 = \mu - 0.6745\sigma$ and $Q_3 = \mu + 0.6745\sigma$. **(c)** For human pregnancies, $Q_1 = 266 - 0.6745 \times 16 \doteq 255.2$ and $Q_3 = 266 + 0.67455 \times 16 \doteq 276.8$ days.

1.147. **(a)** As the quartiles for a standard Normal distribution are ± 0.6745, we have $IQR = 1.3490$. **(b)** $c = 1.3490$: For a $N(\mu, \sigma)$ distribution, the quartiles are $Q_1 = \mu - 0.6745\sigma$ and $Q_3 = \mu + 0.6745\sigma$.

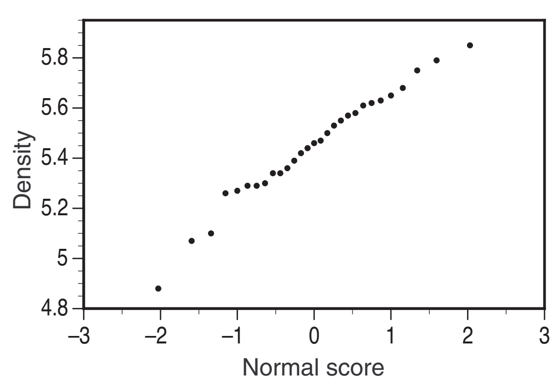
1.148. In the previous two exercises, we found that for a $N(\mu, \sigma)$ distribution, $Q_1 = \mu - 0.6745\sigma$, $Q_3 = \mu + 0.6745\sigma$, and $IQR = 1.3490\sigma$. Therefore, $1.5 \times IQR = 2.0235\sigma$, and the suspected outliers are below $Q_1 - 1.5 \times IQR = \mu - 2.698\sigma$, and above $Q_3 + 1.5 \times IQR = \mu + 2.698\sigma$. The percentage outside of this range is $2 \times 0.0035 = 0.70\%$.

1.149. **(a)** The first and last deciles for a standard Normal distribution are ± 1.2816. **(b)** For a $N(9.12, 0.15)$ distribution, the first and last deciles are $\mu - 1.2816\sigma \doteq 8.93$ and $\mu + 1.2816\sigma \doteq 9.31$ ounces.

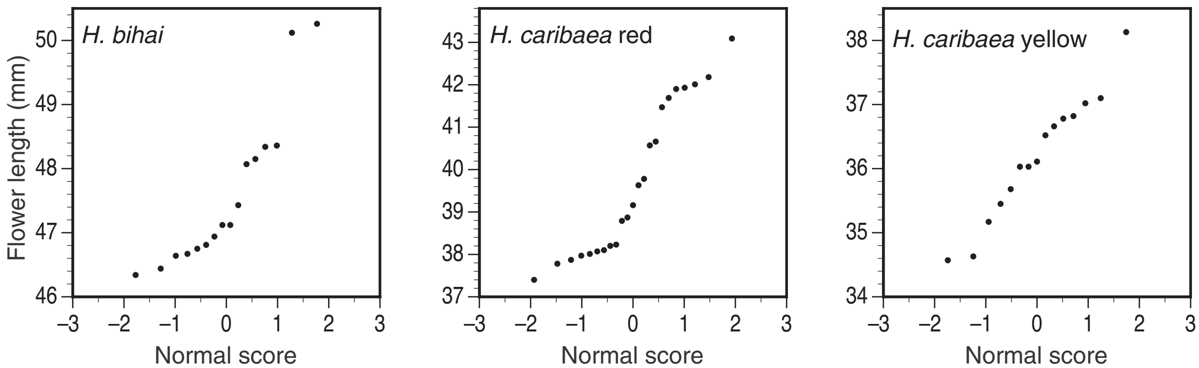
1.150. The shape of the quantile plot suggests that the data are right-skewed (as was observed in Exercises 1.36 and 1.74). This can be seen in the flat section in the lower left—these numbers were less spread out than they should be for Normal data—and the three apparent outliers (the United States, Canada, and Australia) that deviate from the line in the upper right; these were much larger than they would be for a Normal distribution.

1.151. **(a)** The plot is reasonably linear except for the point in the upper right, so this distribution is roughly Normal, but with a high outlier. **(b)** The plot is fairly linear, so the distribution is roughly Normal. **(c)** The plot curves up to the right—that is, the large values of this distribution are larger than they would be in a Normal distribution—so the distribution is skewed to the right.

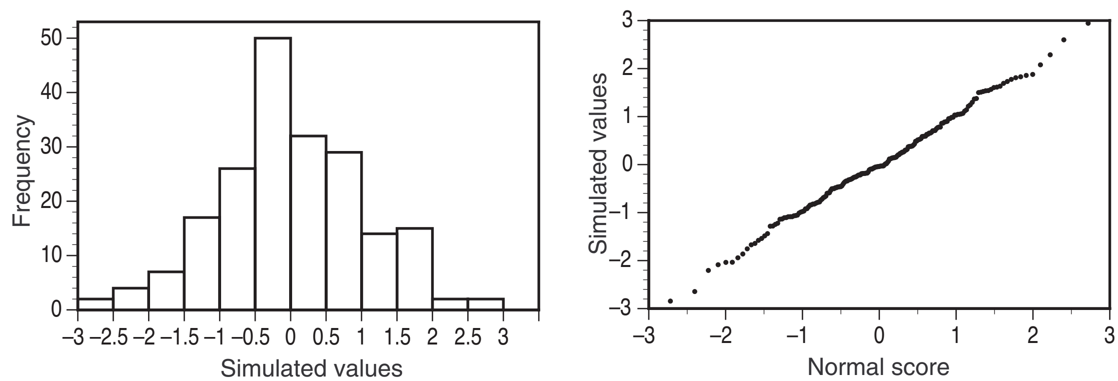
1.152. See also the solution to Exercise 1.42. The plot suggests no major deviations from Normality, although the three lowest measurements do not quite fall in line with the other points.



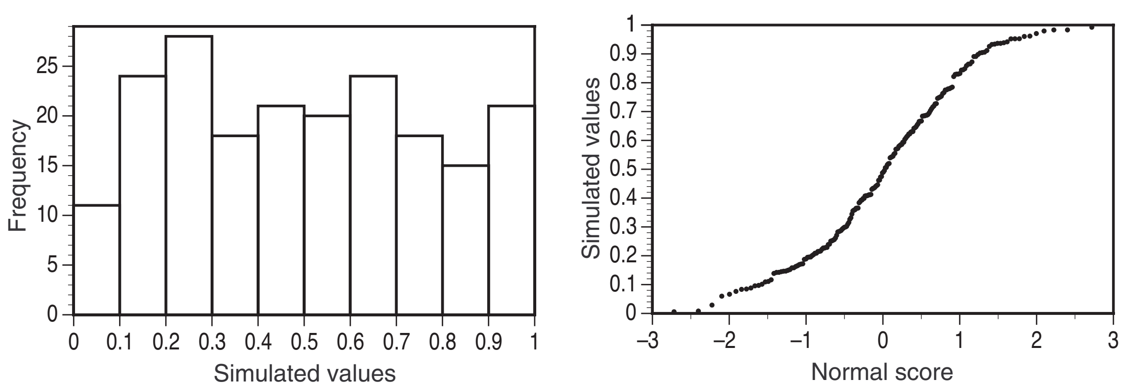
1.153. (a) All three quantile plots are below; the yellow variety is the nearest to a straight line. **(b)** The other two distributions are slightly right-skewed (the lower-left portion of the graph is somewhat flat); additionally, the *bihai* variety appears to have a couple of high outliers.



1.154. Shown are a histogram and quantile plot for one sample of 200 simulated $N(0, 1)$ points. Histograms will vary slightly but should suggest a bell curve. The Normal quantile plot shows something fairly close to a line but illustrates that, even for actual Normal data, the tails may deviate slightly from a line.



1.155. Shown are a histogram and quantile plot for one sample of 200 simulated uniform data points. Histograms will vary slightly but should suggest the density curve of Figure 1.34 (but with more variation than students might expect). The Normal quantile plot shows that, compared to a Normal distribution, the uniform distribution does not extend as low or as high (not surprising, since all observations are between 0 and 1).



1.156. Shown is a back-to-back stemplot; the distributions could also be compared with histograms or boxplots. Either mean/standard deviation or the five-number summary could be used; both are given below. Both the graphical and numerical descriptions reveal that hatchbacks generally have higher fuel efficiency (and also are more variable).

	$\bar{x}$	s	Min	Q_1	M	Q_3	Max
Hatchback	22.548	3.423	16	20	21.5	25	30
Large sedan	16.571	1.425	13	16	17.0	17	19

Hatchback		Large sedan
	13	00
	14	
	15	00
00	16	00000000
	17	0000000000
0	18	0000
000	19	00
00000000	20	
0000000	21	
00	22	
00	23	
00000	24	
000	25	
0	26	
00000	27	
0	28	
0	29	
0	30	

1.157. **(a)** The distribution appears to be roughly Normal. **(b)** One could justify using either the mean and standard deviation or the five-number summary:

$\bar{x}$	s	Min	Q_1	M	Q_3	Max
15.27%	3.118%	8.2%	13%	15.5%	17.6%	22.8%

(c) For example, binge drinking rates are typically 10% to 20%. Which states are high, and which are low? One might also note the geographical distribution of states with high binge-drinking rates: The top six states (Wisconsin, North Dakota, Iowa, Minnesota, Illinois, and Nebraska) are all adjacent to one another.

8	28
9	
10	58
11	34
12	023689
13	015788
14	0077
15	13466889
16	01567
17	45677789
18	8
19	148
20	2
21	6
22	8

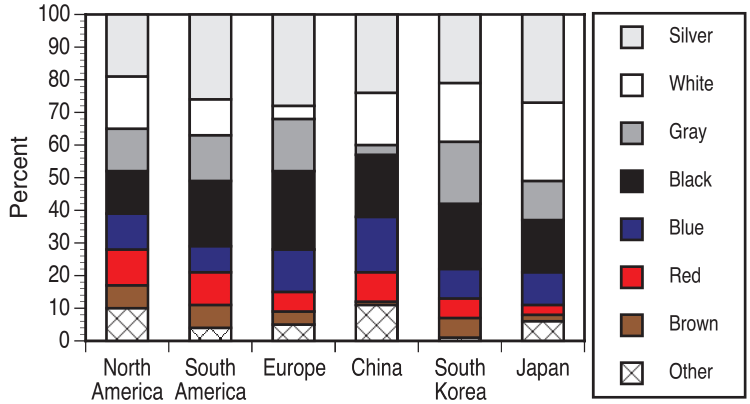
1.158. **(a)** The stemplot on the right suggests that there are two groups of states: the under-23% and over-23% groups. Additionally, while they do not qualify as outliers, Oklahoma (16.3%) and Vermont (30%) stand out as notably low and high. **(b)** One could justify using either the mean and standard deviation or the five-number summary:

$\bar{x}$	s	Min	Q_1	M	Q_3	Max
23.71%	3.517%	16.3%	20.8%	24.3%	26.4%	30%

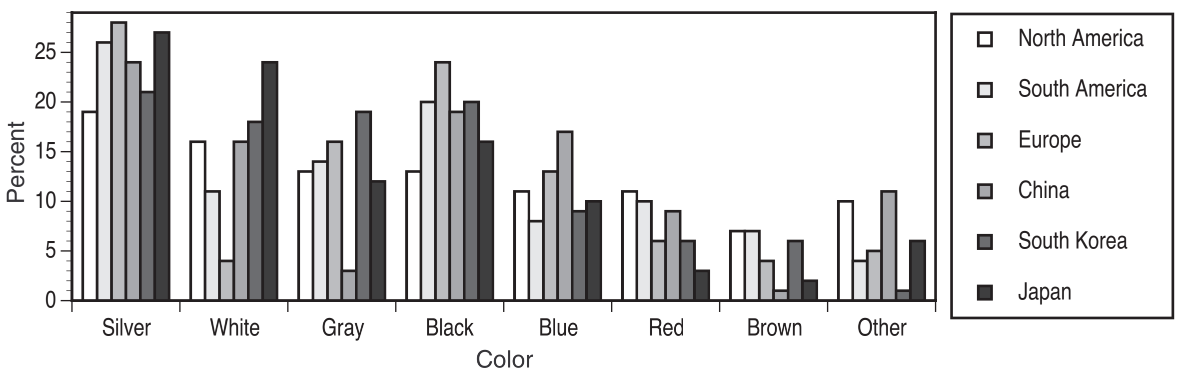
Neither summary reveals the two groups of states visible in the stemplot. **(c)** One could explore the connections (geographical, socioeconomic, etc.) between the states in the two groups; for example, the top group includes many northeastern states, while the bottom group includes quite a few southern states.

16	3
17	
18	14678
19	4679
20	268
21	346899
22	3488
23	
24	12446
25	023468
26	02346
27	0455
28	355679
29	
30	0

1.159. Students might compare color preferences using a stacked bar graph like that shown on the right, or side-by-side bars like those below. (They could also make six pie charts, but comparing slices across pies is difficult.) Possible observations: white is considerably less popular in Europe, and gray is less common in China.



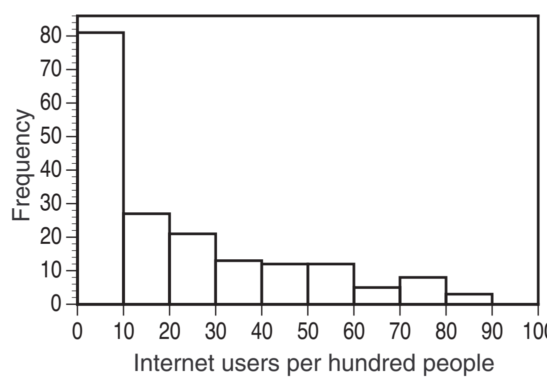
Note: *The orders of countries and colors is as given in the text, which is more-or-less arbitrary. (Colors are ordered by decreasing popularity in North America.)*



1.162. Using either a histogram or stemplot, we see that this distribution is sharply right-skewed. For this reason, the five-number summary is preferred.

Min	Q_1	M	Q_3	Max
0	3	12.5	34	86

Some students might report the less-appropriate $\bar{x} \doteq 21.62$ and $s \doteq 22.76$.



From the histogram and five-number summary, we can observe, for example, that many countries have fewer than 10 Internet users per 100 people. In 75% of countries, less than 1/3 of the population uses the Internet.

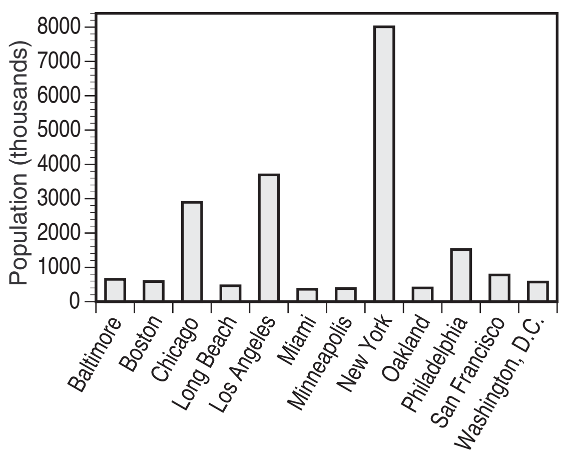
1.163. The distribution is somewhat right-skewed (although considerably less than the distribution with all countries) with only one country (Bosnia and Herzegovina) in the 20's. Because of the irregular shape, students might choose either the mean/standard deviation or the five-number summary:

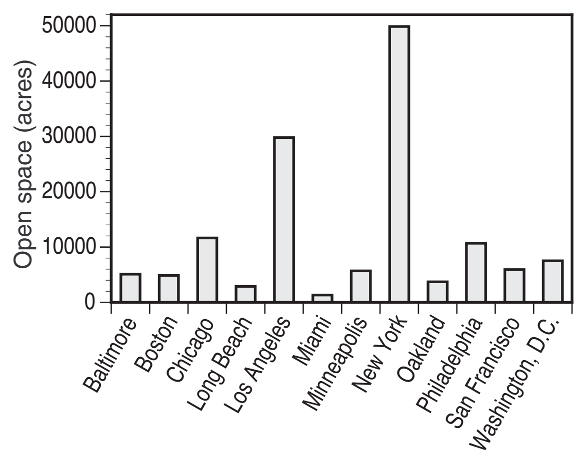
$\bar{x}$	s	Min	Q_1	M	Q_3	Max
39.85	22.05	1.32	18.68	43.185	54.94	85.65

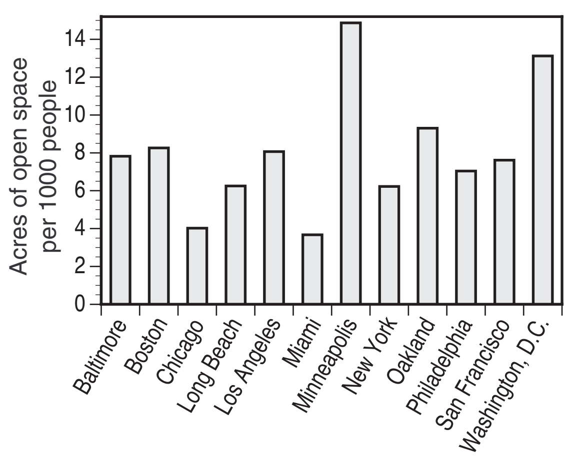
0	145789
1	23488889
2	5
3	0134467
4	124666669
5	022345688
6	223
7	026
8	15

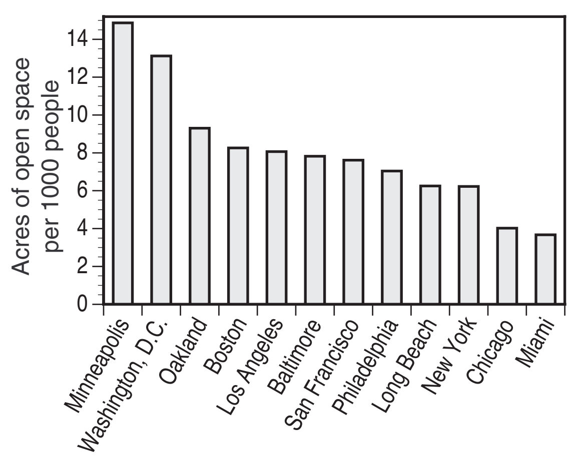
1.164. **(a)** & **(b)** The graphs are below. Bars are shown in alphabetical order by city name (as the data were given in the table). **(c)** For Baltimore, for example, this rate is $\frac{5091}{651} \doteq 7.82$. The complete table is shown on the right. **(d)** & **(e)** Graphs below. Note that the text does not specify whether the bars should be ordered by *increasing* or *decreasing* rate. **(f)** Preferences may vary, but the ordered bars make comparisons easier.

Baltimore	7.82
Boston	8.26
Chicago	4.02
Long Beach	6.25
Los Angeles	8.07
Miami	3.67
Minneapolis	14.87
New York	6.23
Oakland	9.30
Philadelphia	7.04
San Francisco	7.61
Washington, D.C.	13.12









1.165. The given description is true on the average, but the curves (and a few calculations) give a more complete picture. For example, a score of about 675 is about the 97.5th percentile for both genders, so the top boys and girls have very similar scores.

1.166. (a) & **(b)** Answers will vary. Definitions might be as simple as "free time," or "time spent doing something other than studying." For part (b), it might be good to encourage students to discuss practical difficulties; for example, if we ask Sally to keep a log of her activities, the time she spends filling it out presumably reduces her available "leisure time."

1.167. Shown is a stemplot; a histogram should look similar to this. This distribution is relatively symmetric apart from one high outlier. Because of the outlier, the five-number summary (in hours) is preferred:

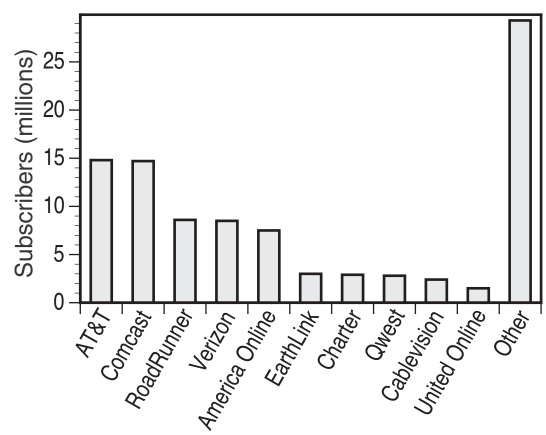
22 23.735 24.31 24.845 28.55

Alternatively, the mean and standard deviation are $\bar{x} = 24.339$ and $s = 0.9239$ hours.

22	013
22	7899
23	000011222233344444
23	55566666667777778888888999
24	00000011111112222222223333333333444444
24	555555666666666777777888888999999
25	00001111233344
25	56666889
26	2
26	56
27	2
27	
28	
28	5

1.168. Gender and automobile preference are categorical; age and household income are quantitative.

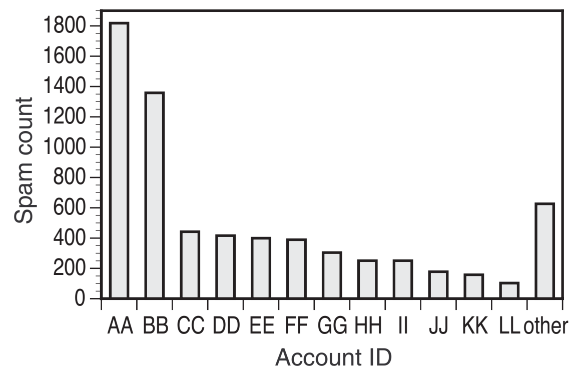
1.169. Either a bar graph or a pie chart could be used. The given numbers sum to 66.7, so the "Other" category presumably includes the remaining 29.3 million subscribers.



1.170. Women's weights are skewed to the right: This makes the mean higher than the median, and it is also revealed in the differences $M - Q_1 = 14.9$ lb and $Q_3 - M = 24.1$ lb.

1.171. (a) For car makes (a categorical variable), use either a bar graph or pie chart. For car age (a quantitative variable), use a histogram, stemplot, or boxplot. **(b)** Study time is quantitative, so use a histogram, stemplot, or boxplot. To show change over time, use a time plot (average hours studied against time). **(c)** Use a bar graph or pie chart to show radio station preferences. **(d)** Use a Normal quantile plot to see whether the measurements follow a Normal distribution.

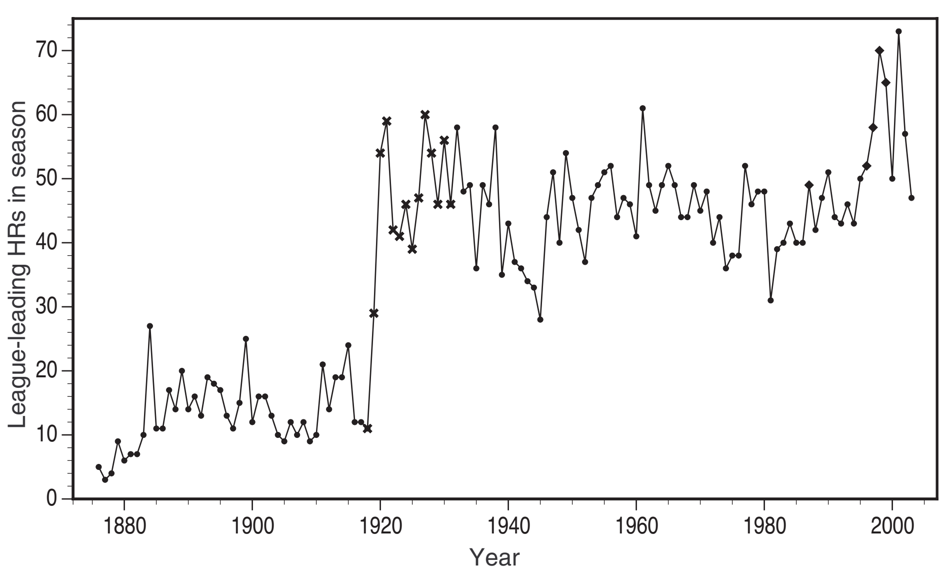
1.172. The counts given add to 6067, so the others received 626 spam messages. Either a bar graph or a pie chart would be appropriate. What students learn from this graph will vary; one observation might be that AA and BB (and perhaps some others) might need some advice on how to reduce the amount of spam they receive.



1.173. No, and no: It is easy to imagine examples of many different data sets with mean 0 and standard deviation 1—for example, {−1,0,1} and {−2,0,0,0,0,0,0,0,2}.

Likewise, for any given five numbers $a \le b \le c \le d \le e$ (not all the same), we can create many data sets with that five-number summary, simply by taking those five numbers and adding some additional numbers in between them, for example (in increasing order): 10, __, 20, __, __, 30, __, __, 40, __, 50. As long as the number in the first blank is between 10 and 20, and so on, the five-number summary will be 10, 20, 30, 40, 50.

1.174. The time plot is shown below; because of the great detail in this plot, it is larger than other plots. Ruth's and McGwire's league-leading years are marked with different symbols. **(a)** During World War II (when many baseball players joined the military), the best home run numbers decline sharply and steadily. **(b)** Ruth seemed to set a new standard for other players; after his first league-leading year, he had 10 seasons much higher than anything that had come before, and home run production has remained near that same level ever since (even the worst post-Ruth year—1945—had more home runs than the best pre-Ruth season). While some might argue that McGwire's numbers also raised the standard, the change is not nearly as striking, nor did McGwire maintain it for as long as Ruth did. (This is not necessarily a criticism of McGwire; it instead reflects that in baseball, as in many other endeavors, rates of improvement tend to decrease over time as we reach the limits of human ability.)



1.175. Bonds's mean changes from 36.56 to 34.41 home runs (a drop of 2.15), while his median changes from 35.5 to 34 home runs (a drop of 1.5). This illustrates that outliers affect the mean more than the median.

1	69
2	4
2	55
3	3344
3	77
4	02
4	5669
5	
5	
6	
6	
7	3

1.176. Recall the text's description of the effects of a linear transformation $x_{\text{new}} = a + bx$: The mean and standard deviation are each multiplied by b (technically, the standard deviation is multiplied by $|b|$, but this problem specifies that $b > 0$). Additionally, we add a to the (new) mean, but a does not affect the standard deviation. **(a)** The desired transformation is $x_{\text{new}} = -40 + 2x$; that is, $a = -40$ and $b = 2$. (We need $b = 2$ to double the standard deviation; as this also doubles the mean, we then subtract 40 to make the new mean 100.) **(b)** $x_{\text{new}} = -45.4545 + 1.8182x$; that is, $a = -49\frac{1}{11} \doteq -49.0909$ and $b = \frac{20}{11} \doteq 1.8182$. (This choice of b makes the new standard deviation 20 and the new mean $145\frac{5}{11}$; we then subtract 45.4545 to make the new mean 100.) **(c)** David's score—$2 \cdot 72 - 40 = 104$—is higher within his class than Nancy's score—$1.8182 \cdot 78 - 45.4545 \doteq 96.4$—is within her class. **(d)** A third-grade score of 75 corresponds to a score of 110 from the $N(100, 20)$ distribution, which has a standard score of $z = \frac{110-100}{20} = 0.5$. (Alternatively, $z = \frac{75-70}{10} = 0.5$.) A sixth-grade score of 75 corresponds to about 90.9 on the transformed scale, which has standard score $z = \frac{90.9-100}{20} = \frac{75-80}{11} \doteq -0.45$. Therefore, about 69% of third graders and 32% of sixth graders score below 75.

1.177. Results will vary. One set of 20 samples gave the results at the right (Normal quantile plots are not shown).

Theoretically, $\bar{x}$ will have a Normal distribution with mean 25 and standard deviation $8/\sqrt{30} \doteq 1.46$, so that about 99.7% of the time, one should find $\bar{x}$ between 20.6 and 29.4. Meanwhile, the theoretical distribution of s is nearly Normal (slightly skewed) with mean $\doteq 7.9313$ and standard deviation $\doteq 1.0458$; about 99.7% of the time, s will be between 4.8 and 11.1.

Means		Standard deviations	
22	568	5	6
23		6	
23	89	6	66899
24	02	7	3
24	89	7	
25	3	8	113
25	6799	8	789
26	124	9	000
26	59	9	556
27	4	10	2

Note: *If we take a sample of size n from a Normal distribution and compute the sample standard deviation S, then* $(S/\sigma)\sqrt{n-1}$ *has a "chi" distribution with* $n-1$ *degrees of freedom (which looks like a Normal distribution when n is reasonably large). You can learn all you would want to know—and more—about this distribution on the Web (for example, at Wikipedia). One implication of this is that "on the average," s underestimates* σ*; specifically, the mean of S is* $\sigma\left(\frac{\sqrt{2}\,\Gamma(n/2)}{\sqrt{n-1}\,\Gamma(n/2-1/2)}\right)$. *The factor in parentheses is always less than 1, but approaches 1 as n approaches infinity. The proof of this fact is left as an exercise—for the instructor, not for the average student!*

Chapter 2 Solutions

2.1. The cases are students.

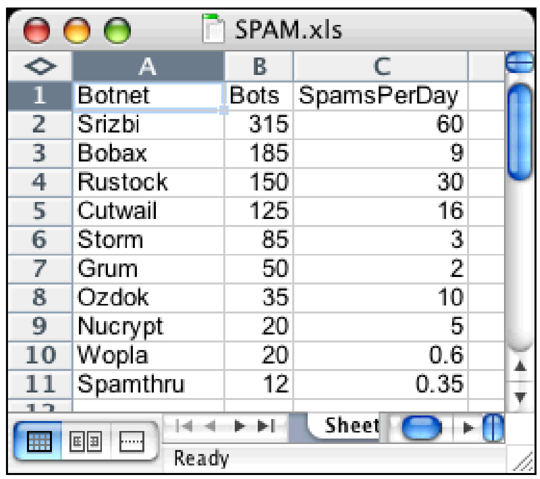
2.2. When students are classified like this, PSQI is being used as a categorical variable, because each student is categorized by the group he/she falls in.

One advantage is that it might simplify the analysis, or at least it might simplify the process of describing the results. (Saying that someone fell into the "poor" category is easier to interpret than saying that person had a PSQI score of 12.) A more subtle issue is that it is not clear whether finding an average is appropriate for these numbers; technically, averages are not appropriate for a quantitative measurement unless the variable is measured on an "interval" scale, meaning (for example) that the difference between PSQI scores of 1 and 2 is the same as the difference between PSQI scores of 10 and 11.

2.3. With this change, the cases are cups of Mocha Frappuccino (as before). The variables (both quantitative) are size and price.

2.4. One could make the argument that being subjected to stress makes it more difficult to sleep, so that SUDS (stress level) is explanatory and PSQI (sleep quality) is the response.

2.5. **(a)** The spreadsheet should look like the image on the right (especially if students use the data file from the companion CD). **(b)** There are 10 cases. **(c)** The image on the right shows the column headings used on the companion CD; some students may create their own spreadsheets and use slightly different headings. (The values of the variables should be the same.) **(d)** The variables in the second and third columns ("Bots" and "SpamsPerDay") are quantitative.


SPAM.xls

	A	B	C
1	Botnet	Bots	SpamsPerDay
2	Srizbi	315	60
3	Bobax	185	9
4	Rustock	150	30
5	Cutwail	125	16
6	Storm	85	3
7	Grum	50	2
8	Ozdok	35	10
9	Nucrypt	20	5
10	Wopla	20	0.6
11	Spamthru	12	0.35

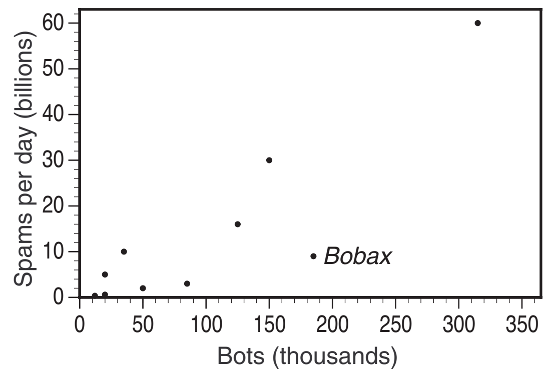
2.6. Stemplots are shown; histograms would be equivalent. Students may choose different ways to summarize the data, such as bar graphs (one bar for each botnet). Note that summarizing each variable separately does not reveal the *relationship* between the two variables; that is done using a scatterplot in the next exercise.

Because both distributions are skewed, we prefer five-number summaries to the mean and standard deviation.

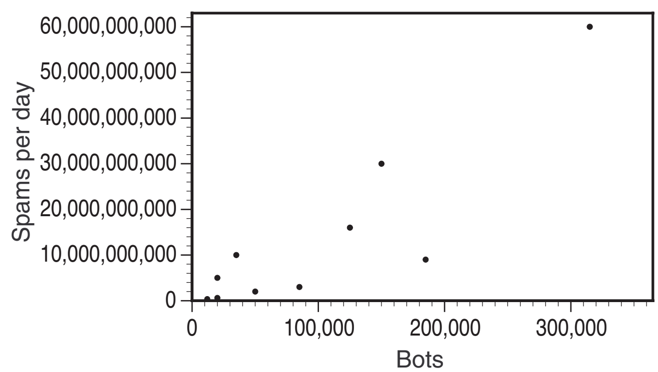
Bots		Spams/day	
0	1223	0	002359
0	58	1	06
1	2	2	
1	58	3	0
2		4	
2		5	
3	1	6	0

	$\bar{x}$	s	Min	Q_1	M	Q_3	Max
Bots (thousands)	99.7	96.6	12	20	67.5	150	315
Spams/day (billions)	13.6	18.6	0.35	2	7.0	16	60

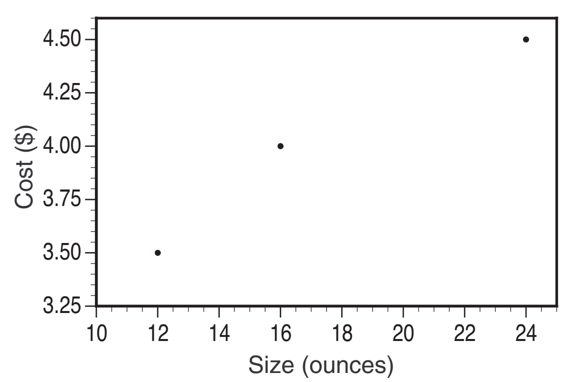
2.7. **(a)** The scatterplot is on the right.
(b) Bobax is the second point from the right. (Bobax has the second-highest bot count with 185 thousand, but is relatively low in spam messages at 9 billion per day.)



2.8. **(a)** The resulting spreadsheet is not shown. **(b)** Scatterplot on the right. **(c)** The points are arranged exactly as before, but the large numbers on the axes are distracting.



2.9. Size seems to be the most reasonable choice for explanatory variable because it seems nearly certain that Starbucks first decided which sizes to offer, then determined the appropriate price for each size (rather than vice versa). The scatterplot shows a positive association between size and price.



2.10. Two good choices are the change in debt from 2006 to 2007 (subtract the two numbers for each country) or the ratio of the two debts (divide one number by the other). Students may think of other new variables, but these have the most direct bearing on the question.

Shown are stemplots of the increase (2007 debt minus 2006 debt, measured in US$ billions), and the debt ratio (2007 debt divided by 2006 debt; these numbers have no units). From either variable, we can see that debt increased for all but two of the 24 countries. This can be summarized using either the mean and standard deviation or the five-number summary (the latter is preferred for increase, because of the skew).

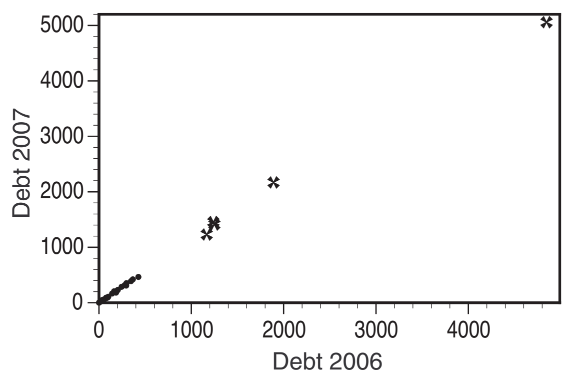
Increase	
−0	21
0	0044
0	56667
1	034
1	
2	13
2	7
3	
3	77
4	122
4	
5	34

Ratio	
9	88
10	4
10	668
11	0223334
11	66778
12	013
12	589

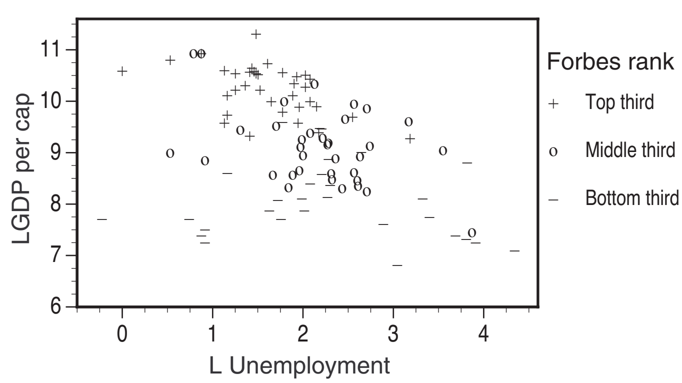
	$\bar{x}$	s	Min	Q_1	M	Q_3	Max
Increase	19.07	18.38	−2.89	5.25	12.205	37.66	54.87
Ratio	1.145	0.082	0.984	1.098	1.143	1.193	1.298

Note: *In looking at increases, one notes that the size of the debt and the size of the change are related (countries with smaller debts typically changed less than countries with large debts). Debt ratio does not have this relationship with debt size (or at least it is less apparent); for this reason, it might be considered a better choice for answering this question.*

2.11. The new points (marked with a different symbol) are far away from the others, but fall roughly in the same line, so the relationship is essentially unchanged: It is still strong, linear, and positive.



2.12. Student choices of symbols will vary; the plot on the right uses +, o, −, rather than the more obvious H, M, L. (The latter symbols are harder to distinguish when overlapping.)

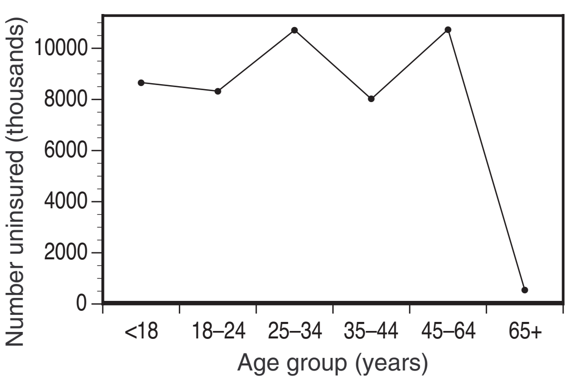


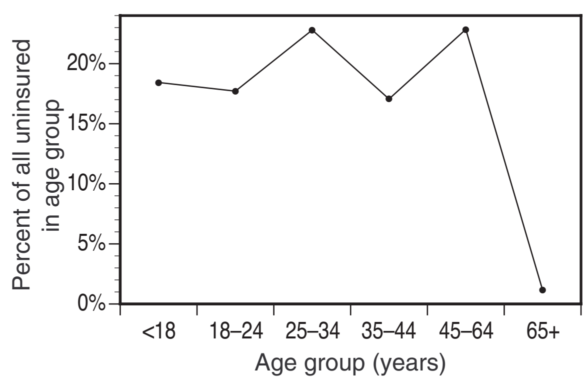
This graph reinforces the observation in Example 2.14 that GDP ties in closely with rankings; generally, high GDP goes with high rank, middle GDP with middle rank, and low GDP with low rank. As before, the relationship (if any) between unemployment and rankings is not so clear.

2.13. **(a)** A boxplot summarizes the distribution of one variable. (Two [or more] boxplots can be used to compare two [or more] distributions, but that does not allow us to examine the relationship between those variables.) **(b)** This is only correct if there is an explanatory/response relationship. Otherwise, the choice of which variable goes on which axis might be somewhat arbitrary. **(c)** High values go with high values, and low values go with low values. (Of course, those statements are generalizations; there can be exceptions.)

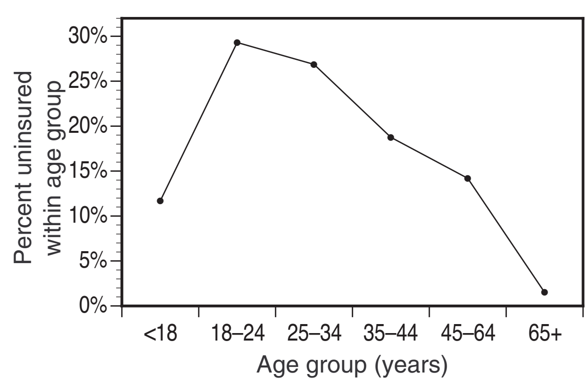
2.14. **(a)** The points should all fall close to a negatively sloped line. **(b)** Look for a "cloud" of points with no discernible pattern. Watch for students who mistakenly consider "no relationship" as meaning "no linear relationship." For example, points that suggest a circle, triangle, or curve may indicate a non-linear relationship. **(c)** The points should be widely scattered around a positively sloped line. **(d)** Sketches might be curved, angular, or something more imaginative.

2.15. **(a)** Below, left. **(b)** Adding up the numbers in the first column of the table gives 46,994 thousand (that is, about 47 million) uninsured; divide each number in the second column by this amount. (Express answers as percents; that is, multiply by 100.) **(c)** Below, right. (The title on the vertical axis is somewhat wordy, but that is needed to distinguish between this graph and the one in the solution to the next exercise.) **(d)** The plots differ only in the vertical scale. **(e)** The uninsured are found in similar numbers for the five lowest age groups (with slightly more in those aged 25–34 and 45–64), and fewer among those over 65.





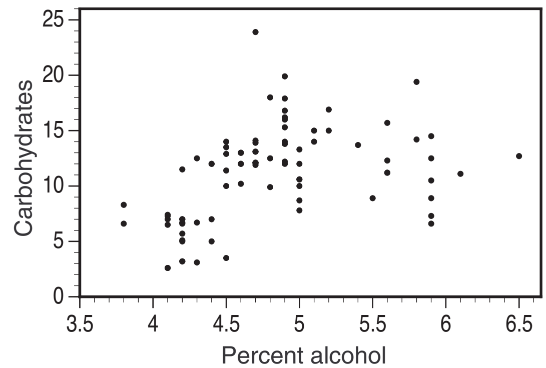
2.16. **(a)** For example, in the under-18 age group, $\frac{8661}{74{,}101} \doteq 0.1169 = 11.69\%$; in the 18–24 age group, $\frac{8323}{28{,}405} \doteq 0.2930 = 29.30\%$; etc. **(b)** Plot on the right. The title on the vertical axis is rather long to distinguish between this graph and the second one in the previous solution. **(c)** The youngest (often covered by their parents' insurance) and oldest (covered by Medicare) are least likely to be uninsured. For the age groups in between, the percent uninsured gradually decreases.



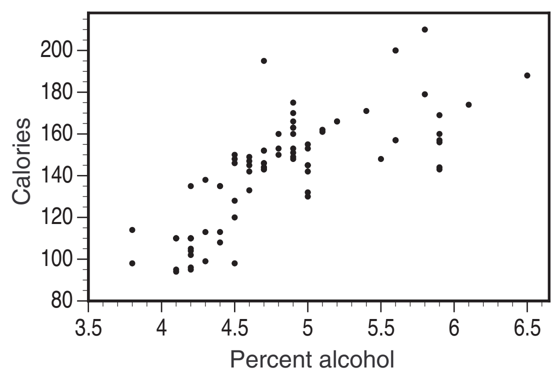
2.17. The percents in Exercise 2.15 show what fraction of the uninsured fall in each age group. The percents in Exercise 2.16 show what fraction of each age group is uninsured.

Note: *When looking at fractions and percents, encourage students to focus on the "whole"—that is, what does the denominator represent? For all the fractions in Exercise 2.15, the "whole" is the group of all uninsured people. For Exercise 2.16, the "whole" for each fraction is the total number of people in the corresponding age group.*

2.18. See also the solutions to Exercises 1.62 and 1.64. **(a)** Apart from the outlier, the scatterplot suggests a weak positive relationship—that is, beers with high alcohol content generally have more carbohydrates, and those with low alcohol content generally have fewer carbohydrates. **(b)** The outlier is O'Doul's, which is marketed as "non-alcoholic" beer. **(c)** The scatterplot is on the right. **(d)** Without the outlier, the scatterplot suggests a slightly curved relationship. (That relationship was also somewhat visible in Figure 2.10, but it is easier to see when the points are not crowded together in half of the graph.)



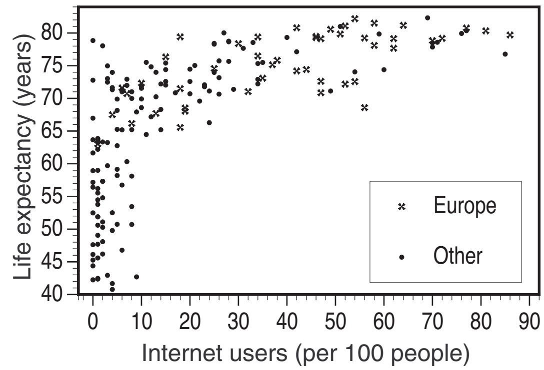
2.19. **(a)** The scatterplot is on the right. **(b)** There is a moderate positive linear relationship. (There is some suggestion of a curve, but a line seems to be a reasonable approximation given the amount of scatter.)



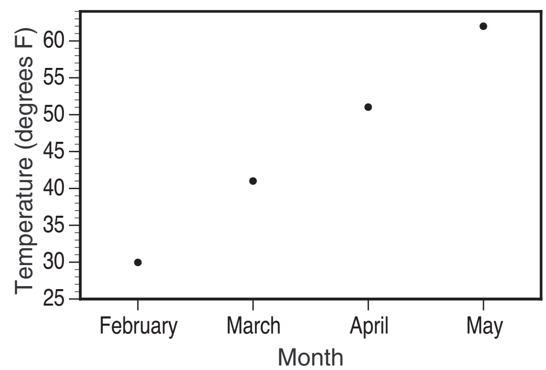
2.20. **(a)** Figure 2.11 shows a positive curved relationship. More specifically, in countries with fewer than 10 Internet users per 100 people, life expectancy ranges between 40 and about 75 years. For countries with more than 10 Internet users per 100 people, life expectancy increases (slowly) with increasing Internet usage. **(b)** A more likely explanation for the association is that countries with higher Internet usage are typically more developed and affluent, which comes with benefits such as better access to medical care, etc.

2.21. There is a moderate positive linear relationship; the relationship for all countries is less linear because of the wide range in life expectancy among countries with low Internet use.

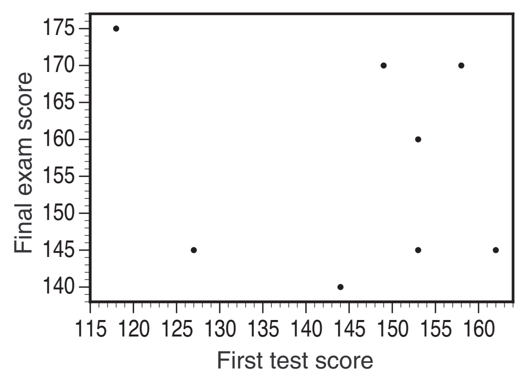
2.22. Students might choose different ways of creating a single plot; the most obvious choice is to use different symbols for European and other countries. (This requires students either to separate the European countries from the others, or to superimpose one graph on another.)



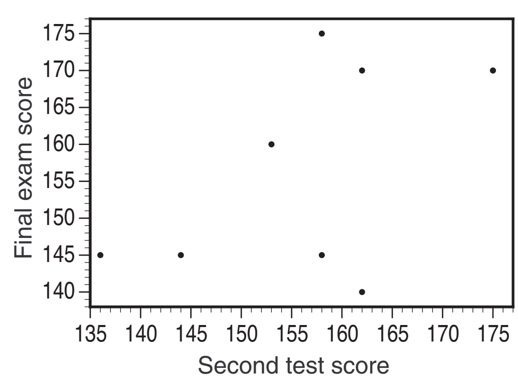
2.23. (a) "Month" (the passage of time) explains changes in temperature (not vice versa). **(b)** Temperature increases linearly with time (about 10 degrees per month); the relationship is strong.



2.24. (a) First test score should be explanatory because it comes first chronologically. **(b)** The scatterplot shows no clear association; however, the removal of one point (the sixth student, in the upper left corner of the scatterplot) leaves a weak-to-moderate positive association. **(c)** A few students can disrupt the pattern quite a bit; for example, perhaps the sixth student studied very hard after scoring so low on the first test, while some of those who did extremely well on the first exam became overconfident and did not study hard enough for the final (the points in the lower right corner of the scatterplot).



2.25. (a) The second test happens before the final exam, so that score should be viewed as explanatory. **(b)** The scatterplot shows a weak positive association. **(c)** Students' study habits are more established by the middle of the term.



2.26. To be considered an outlier, the point for the ninth student should be in either the upper left or lower right portion of the scatterplot. The former would correspond to a student who had a below-average second-test score but an above-average final-exam score. The latter would be a student who did well on the second test but poorly on the final.

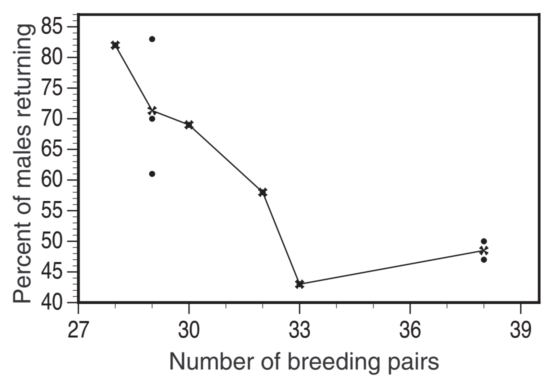
2.27. **(a)** Age is explanatory; weight is the response variable. **(b)** Explore the relationship; there is no reason to view one or the other as explanatory. **(c)** Number of bedrooms is explanatory; price is the response variable. **(d)** Amount of sugar is explanatory; sweetness is the response variable. **(e)** Explore the relationship.

2.28. Parents' income is explanatory, and college debt is the response. Both variables are quantitative. We would expect a negative association: Low income goes with high debt, high income with low debt.

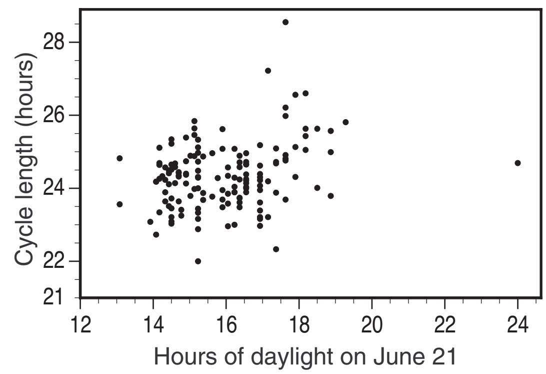
2.29. **(a)** In general, we expect more intelligent children to be better readers and less intelligent children to be weaker. The plot does show this positive association. **(b)** The four points are for children who have moderate IQs but poor reading scores. **(c)** The rest of the scatterplot is roughly linear but quite weak (there would be a lot of variation about any line we draw through the scatterplot).

2.30. **(a)** The response variable (estimated level) can only take on the values 1, 2, 3, 4, 5, so the points in the scatterplot must fall on one of those five levels. **(b)** The association is (weakly) positive. **(c)** The estimate is 4, which is an overestimate; that child had the lowest score on the test.

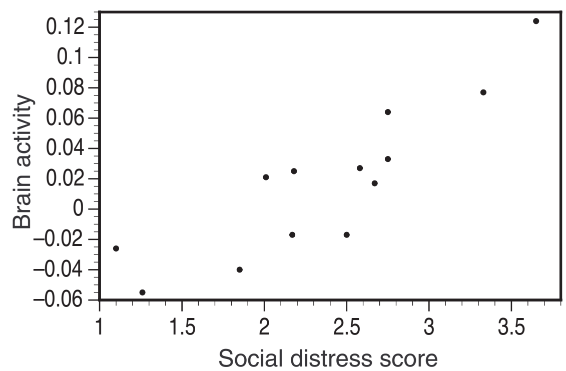
2.31. **(a)** If we used the number of males returning, then we might not see the relationship because areas with many breeding pairs would correspondingly have more males that might potentially return. (In the given numbers, the number of breeding pairs varies only from 28 to 38, but considering hypothetical data with 10 and 100 breeding pairs makes more apparent the reason for using percents rather than counts.) **(b)** Scatterplot on the right. Mean responses are shown as crosses; the mean responses with 29 and 38 breeding pairs are (respectively) 71.3333% and 48.5% males returning. **(c)** The scatterplot does show the negative association we would expect if the theory were correct.



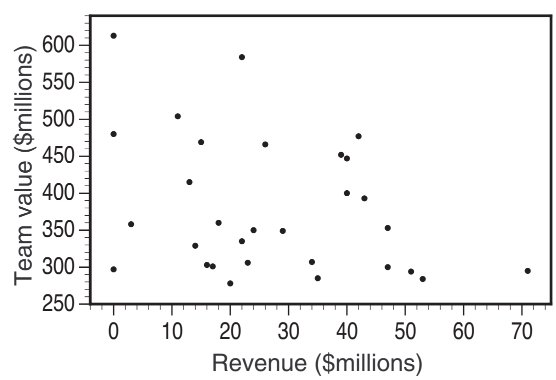
2.32. There appears to be a positive association between cycle length and day length, but it is quite weak: The points of the scatterplot are generally located along a positively sloped line but with a lot of spread around that line. (Ideally, both axes should have the same scale.)

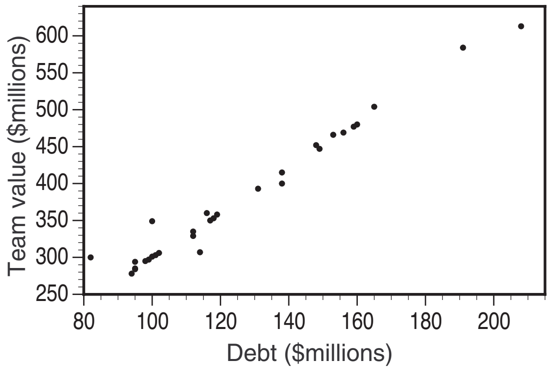


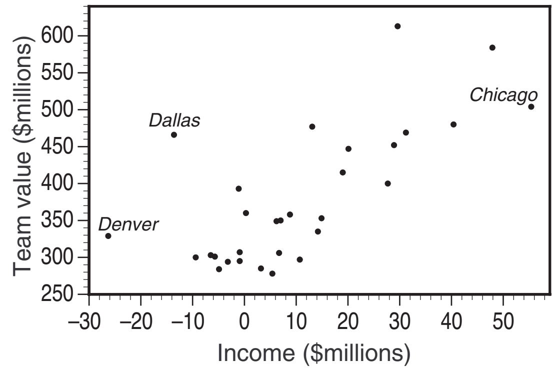
2.33. The scatterplot shows a fairly strong, positive, linear association. There are no particular outliers; each variable has low and high values, but those points do not deviate from the pattern of the rest. Social exclusion does appear to trigger a pain response.



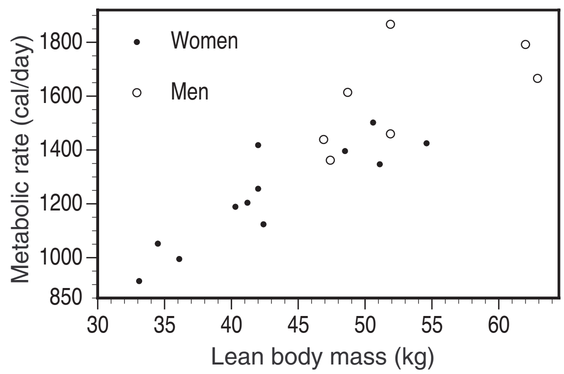
2.34. **(a)** Value and revenue (scatterplot on the right) have a weak negative relationship. The form of the relationship is unclear; one would hesitate to call it linear. **(b)** Value and debt (scatterplot following page, left) have a very strong, positive, linear association. **(c)** Value and income (scatterplot following page, right) have a positive, roughly linear relationship. The relationship is quite weak, partly because the two teams that lost the most money (Denver and Dallas) had higher-than-expected values, and the team with the most income (Chicago) had a lower-than-expected value. Without those three points, we might call the relationship moderately strong. **(d)** Debt has the strongest association with value; other observations will vary.



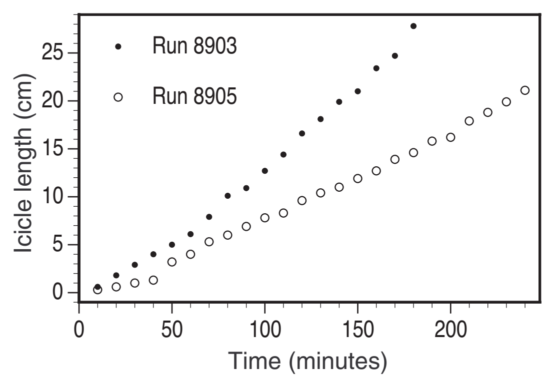




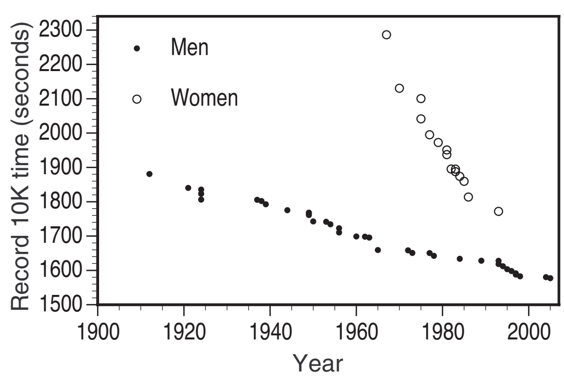
2.35. **(a)** Women are marked with filled circles, men with open circles. **(b)** The association is linear and positive. The women's points show a stronger association. As a group, males typically have larger values for both variables.



2.36. **(a)** In the scatterplot on the right, the open circles represent run 8905, the higher flow rate. **(b)** Icicles seem to grow faster when the water runs more slowly. (Note that there is no guarantee that the pattern we observe with these two flow rates applies to rates a lot faster than 29.6 mg/s, or slower than 11.9 mg/s.)



2.37. **(a)** Both men (filled circles) and women (open circles) show fairly steady improvement. Women have made more rapid progress, but their progress seems to have slowed (the record has not changed since 1993), while men's records may be dropping more rapidly in recent years. **(b)** The data support the first claim but do not seem to support the second.



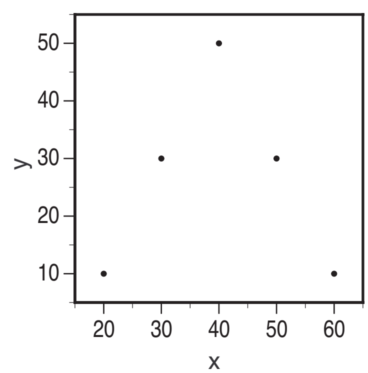
2.38. The correlation is $r \doteq 0.8839$. (This can be computed by hand, but software makes it much easier.)

2.39. **(a)** The correlation is $r \doteq 0.8839$ (again). **(b)** They are equal. **(c)** Units do not affect correlation.

2.40. The correlation is near 1, because the scatterplot shows a very strong positive linear association. (This can be confirmed with the data; we find that $r \doteq 0.9971$.)

2.41. In both these cases, the points in a scatterplot would fall exactly on a positively sloped line, so both have correlation $r = 1$. **(a)** With $x =$ the price of a brand-name product, and $y =$ the store-brand price, the prices satisfy $y = 0.9x$. **(b)** The prices satisfy $y = x - 1$.

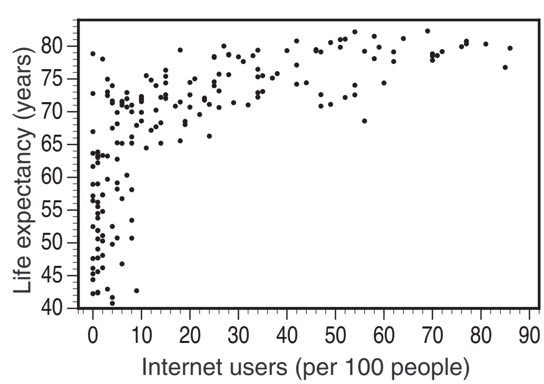
2.42. **(a)** Scatterplot on the right. **(b)** The relationship is very strong (assuming these five points are truly representative of the overall pattern), but it is not linear. **(c)** The correlation is $r = 0$. **(d)** Correlation only measures the strength of *linear* relationships.



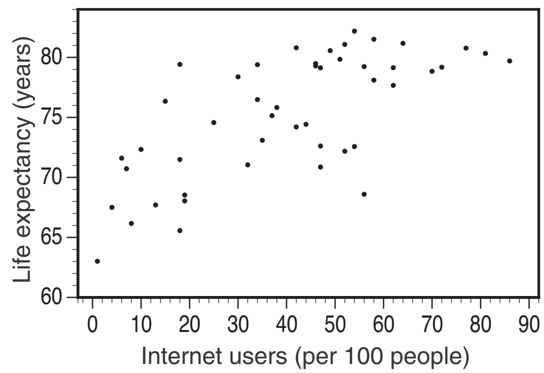
2.43. Software gives $r \doteq 0.2873$. (This is consistent with the not-very-strong association visible in the plot.)

2.44. **(a)** With the outlier (O'Doul's) removed, the correlation is $r \doteq 0.4185$. **(b)** Outliers that are not consistent with the pattern of the other points tend to decrease the size of r (that is, they make r closer to 0), because they weaken the association between the variables. (Points that lie far from the other points, but roughly on the same line, are typically not referred to as outliers because they actually *strengthen* the association.)

2.45. The correlation is $r \doteq 0.6701$, but we note that because this relationship is not linear, r is not really appropriate for this situation.



2.46. **(a)** Scatterplot on the right. **(b)** The correlation is $r \doteq 0.6862$. **(c)** The relationship (as measured by r) is stronger for European countries than for the larger data set—and more importantly, r is more appropriate for the European data set, because the relationship is at least roughly linear.



2.47. **(a)** $r \doteq 0.5194$. **(b)** The first-test/final-exam correlation will be lower, because the relationship is weaker. (See the next solution for confirmation.)

2.48. **(a)** $r \doteq -0.2013$. **(b)** The small correlation (that is, close to 0) is consistent with a weak association. **(c)** This correlation is much smaller (in absolute value) than the second-test/final-exam correlation 0.5194.

2.49. Such a point should be at the lower left part of the scatterplot. Because it tends to strengthen the relationship, the correlation increases.

Note: *In this case, r was positive, so strengthening the relationship means r gets* larger. *If r had been negative, strengthening the relationship would have* decreased *r (toward −1).*

2.50. Any outlier should make r closer to 0, because it weakens the relationship. To be considered an outlier, the point for the ninth student should be in either the upper left or lower right portion of the scatterplot. The former would correspond to a student who had a below-average second-test score but an above-average final-exam score. The latter would be a student who did well on the second test but poorly on the final.

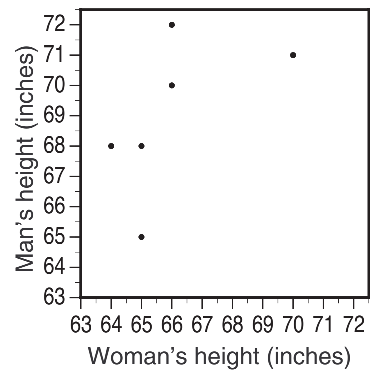
Note: *In this case, because r > 0, this means r gets* smaller. *If r had been negative, getting closer to 0 would mean that r gets* larger *(but gets smaller in absolute value).*

2.51. The correlations are listed on the right; these support the observation from the solution to Exercise 2.34 that the value/debt relationship is by far the strongest.

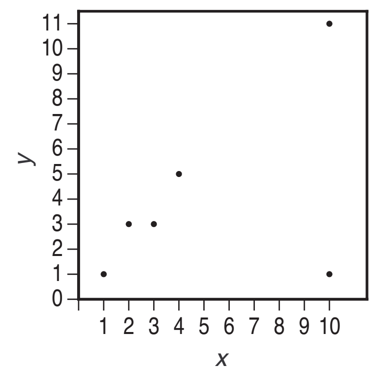
Value and revenue	$r_1 \doteq -0.3228$
Value and debt	$r_2 \doteq 0.9858$
Value and income	$r_3 \doteq 0.7177$

2.52. For Exercise 2.32, $r_1 \doteq 0.2797$; for Exercise 2.36, $r_2 \doteq 0.9958$ (run 8903) and $r_3 \doteq 0.9982$ (run 8905).

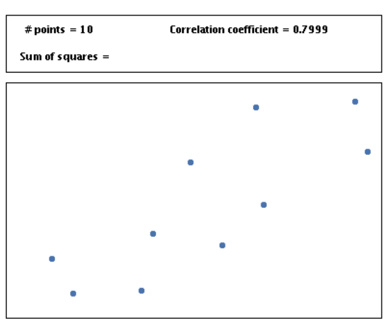
2.53. **(a)** The scatterplot shows a moderate positive association, so r should be positive, but not close to 1. **(b)** The correlation is $r \doteq 0.5653$. **(c)** r would not change if all the men were six inches shorter. A positive correlation does not tell us that the men were generally taller than the women; instead it indicates that women who are taller (shorter) than the average woman tend to date men who are also taller (shorter) than the average man. **(d)** r would not change because it is unaffected by units. **(e)** r would be 1 because the points of the scatterplot would fall exactly on a positively sloped line (with no scatter).

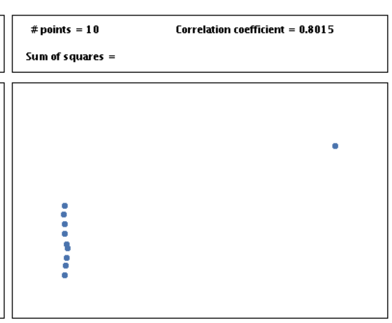


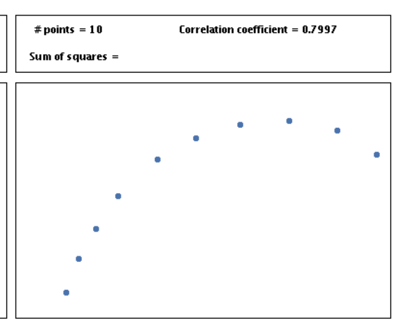
2.54. The correlation is $r \doteq 0.481$. The correlation is greatly lowered by the one outlier. Outliers tend to have fairly strong effects on correlation; it is even stronger here because there are so few observations.



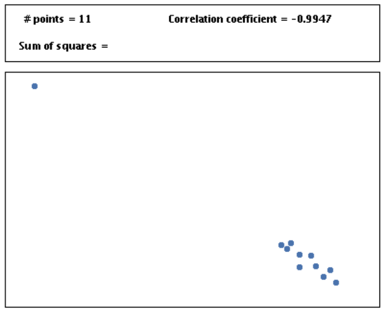
2.55. **(a)** As two points determine a line, the correlation is always either -1 or 1. **(b)** Sketches will vary; an example is shown as the first graph below. Note that the scatterplot must be positively sloped, but r is affected only by the scatter about a line drawn through the data points, not by the steepness of the slope. **(c)** The first nine points cannot be spread from the top to the bottom of the graph because in such a case the correlation cannot exceed about 0.66 (based on empirical evidence—that is, from a reasonable amount of playing around with the applet). One possibility is shown as the second graph below. **(d)** To have $r \doteq 0.8$, the curve must be higher at the right than at the left. One possibility is shown as the third graph below.

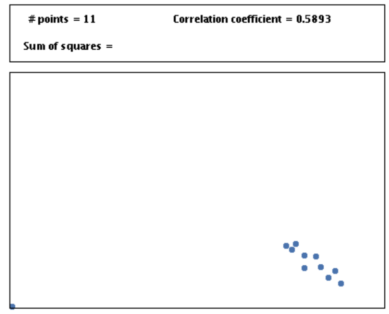






2.56. **(a)** The correlation will be closer to -1. One possible answer is shown below, left. **(b)** Answers will vary, but the correlation will increase and can be made positive by dragging the point down far enough (below, right).





2.57. (Scatterplot not shown.) If the husband's age is y and the wife's x, the linear relationship $y = x + 2$ would hold, and hence $r = 1$ (because the slope is positive).

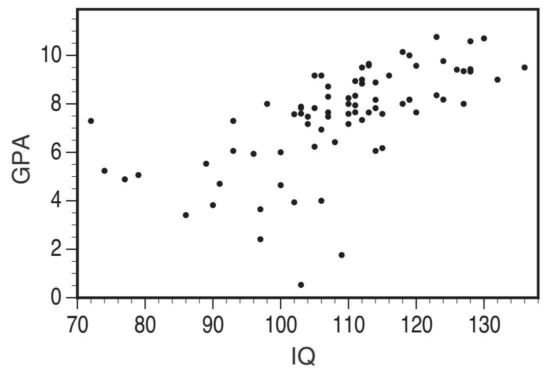
2.58. Explanations and sketches will vary, but should note that correlation measures the strength of the association, not the slope of the line (except for the sign of the slope—positive or negative). The hypothetical Funds A and B mentioned in the report, for example, might be related by a linear formula with slope 2 (or 1/2).

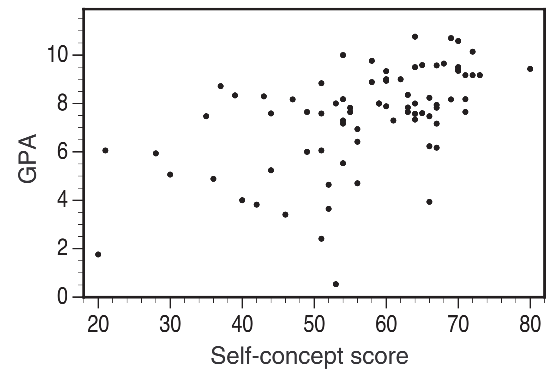
2.59. The person who wrote the article interpreted a correlation close to 0 as if it were a correlation close to -1 (implying a negative association between teaching ability and research productivity). Professor McDaniel's findings mean there is little linear association between research and teaching—for example, knowing that a professor is a good researcher gives little information about whether she is a good or bad teacher.

Note: *Students often think that "negative association" and "no association" mean the same thing. This exercise provides a good illustration of the difference between these terms.*

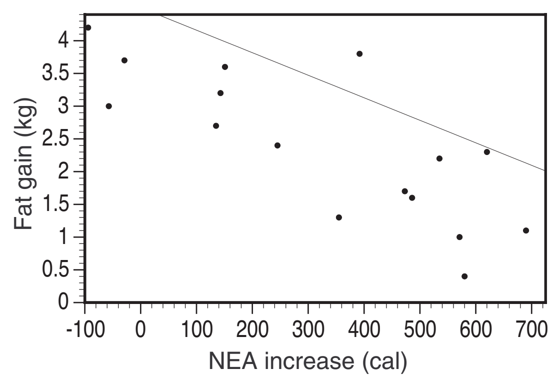
2.60. **(a)** Because occupation has a categorical (nominal) scale, we cannot compute the correlation between occupation and anything. (There may be a strong *association* between these variables; some writers and speakers use "correlation" as a synonym for "association." It is much better to retain the more specific meaning.) **(b)** A correlation $r = 1.19$ is impossible because $-1 \leq r \leq 1$ always. **(c)** Neither variable (gender and color) is quantitative.

2.61. Both relationships (scatterplots follow) are somewhat linear. The GPA/IQ scatterplot ($r \doteq 0.6337$) shows a stronger association than GPA/self-concept ($r \doteq 0.5418$). The two students with the lowest GPAs stand out in both plots; a few others stand out in at least one plot. Generally speaking, removing these points raises r (because the remaining points look more linear). An exception: Removing the lower-left point in the self-concept plot decreases r because the relative scatter of the remaining points is greater.





2.62. The line lies almost entirely above the points in the scatterplot. (The slope -0.00344 of this line is the same as the regression equation given in Example 2.19, but the intercept 4.505 is one more than the regression intercept.)



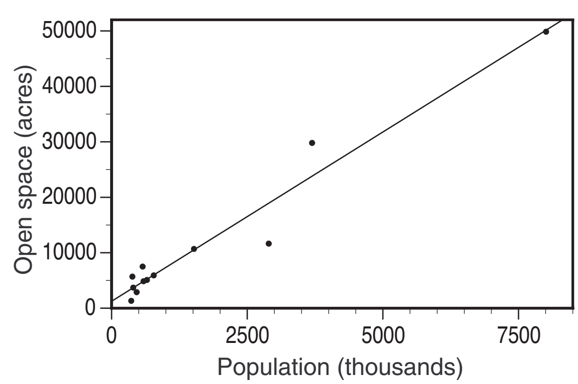
2.63. The estimated fat gain is $3.505 - 0.00344 \times 600 \doteq 1.441$ kg.

2.64. The data used to determine the regression line had NEA increase values ranging from -94 to 690 calories, so estimates for values inside that range (like 200 and 500) should be relatively safe. For values far outside this range (like -400 and 1000), the predictions would not be trustworthy.

2.65. The table on the right shows the values of r^2 (expressed as percents). We observe that (i) the fraction of variation explained depends only on the magnitude (absolute value) of r, not its sign, and (ii) the fraction of explained variation drops off drastically as $|r|$ moves away from 1.

r	-0.9	-0.5	-0.3	0	0.3	0.5	0.9
r^2	81%	25%	9%	0%	9%	25%	81%

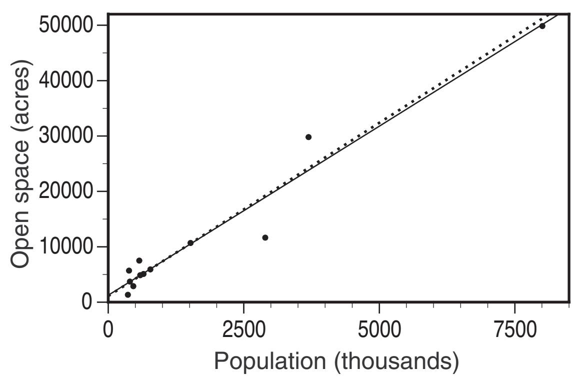
2.66. (a) Scatterplot on the right. **(b)** The association appears to be roughly linear (although note that the slope of the line is almost completely determined by the largest cities). **(c)** The regression equation is $\hat{y} = 1248 + 6.1050x$. **(d)** Regression on population explains $r^2 \doteq 95.2\%$ of the variation in open space.



2.67. Residuals (found with software) are given in the table on the right. Los Angeles is the best; it has nearly 6000 acres more than the regression line predicts. Chicago, which falls almost 7300 acres short of the regression prediction, is the worst of this group.

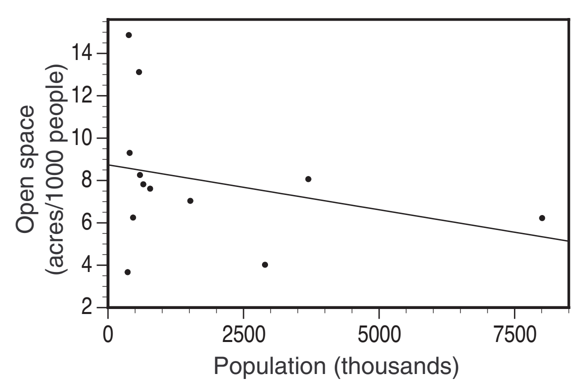
Los Angeles	5994.85
Washington, D.C.	2763.75
Minneapolis	2107.59
Philadelphia	169.42
Oakland	27.91
Boston	20.96
San Francisco	−75.78
Baltimore	−131.55
New York	−282.99
Long Beach	−1181.70
Miami	−2129.21
Chicago	−7283.26

2.68. Because New York's data point is consistent with the pattern of the other cities, we don't consider it an outlier. It does have some impact on the regression line; with New York removed, the equation is $\hat{y} = 1105 + 6.2557x$. However, in the plot on the right, we note that the original regression line (solid) and the new line (dashed) are very similar, and the residuals are likewise very similar.



2.69. For Baltimore, for example, this rate is $\frac{5091}{651} \doteq 7.82$. The complete table is shown below on the left. Note that population is in thousands, so these are in units of acres per 1000 people. **(a)** Scatterplot below on the right. **(b)** The association is much less linear than in the scatterplot for Exercise 2.66. **(c)** The regression equation is $\hat{y} = 8.739 - 0.000424x$. **(d)** Regression on population explains only $r^2 \doteq 8.7\%$ of the variation in open space per person.

Baltimore	7.82
Boston	8.26
Chicago	4.02
Long Beach	6.25
Los Angeles	8.07
Miami	3.67
Minneapolis	14.87
New York	6.23
Oakland	9.30
Philadelphia	7.04
San Francisco	7.61
Washington, D.C.	13.12



2.70. As in Exercise 2.67, we compute residuals to assess whether a city has more or less open space than we would expect. These are given on the right, in descending order. This time, Minneapolis is best, with about 6.3 acres per 1000 people above what we predict. Miami is worst by this measure, falling short of the prediction by almost 5 acres per 1000 people.

Preferences will vary. One reason to prefer the first approach—apart from the stronger, more linear association—is the negative relationship in the second approach. Why would an individual in a large city need less open space than an individual in a smaller city?

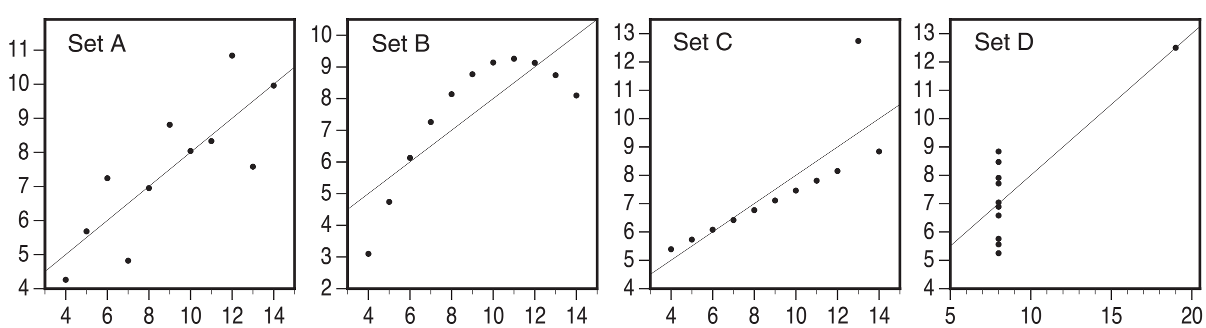
Minneapolis	6.290
Washington, D.C.	4.622
Los Angeles	0.893
New York	0.883
Oakland	0.733
Boston	−0.230
Baltimore	−0.643
San Francisco	−0.796
Philadelphia	−1.056
Long Beach	−2.294
Chicago	−3.490
Miami	−4.914

2.71. The regression equation for predicting carbohydrates from alcohol content is $\hat{y} = 3.379 + 1.6155x$.

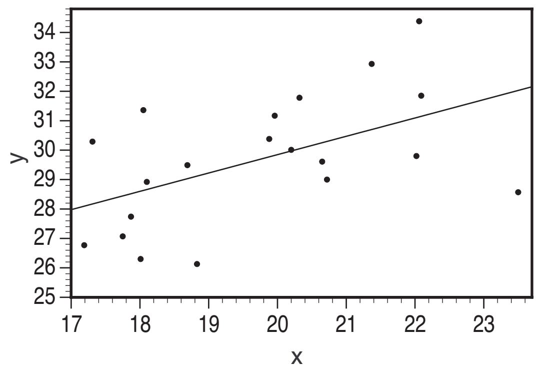
Note: *As we would guess from the scatterplot, and from the correlation* $r \doteq 0.2873$ *found in Exercise 2.43, this is not a very reliable prediction; it only explains* $r^2 \doteq 8.3\%$ *of the variation in carbohydrates.*

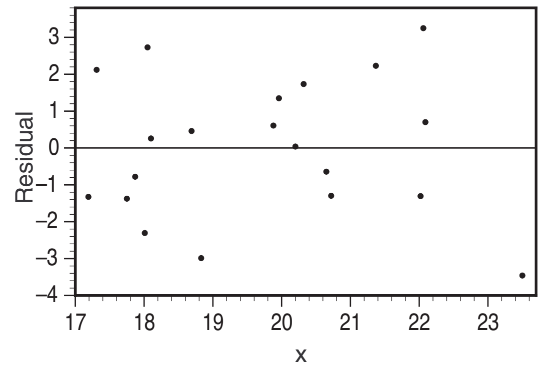
2.72. **(a)** With the outlier (O'Doul's) removed, the regression equation changes to $\hat{y} = -3.544 + 3.0319x$. **(b)** An outlier tends to weaken the association between the variables, and (as in this case) can drastically change the linear regression equation.

2.73. **(a)** To three decimal places, the correlations are all approximately 0.816 (for set D, r actually rounds to 0.817), and the regression lines are all approximately $\hat{y} = 3.000 + 0.500x$. For all four sets, we predict $\hat{y} \doteq 8$ when $x = 10$. **(b)** Scatterplots below. **(c)** For Set A, the use of the regression line seems to be reasonable—the data do seem to have a moderate linear association (albeit with a fair amount of scatter). For Set B, there is an obvious *non*-linear relationship; we should fit a parabola or other curve. For Set C, the point (13, 12.74) deviates from the (highly linear) pattern of the other points; if we can exclude it, the (new) regression formula would be very useful for prediction. For Set D, the data point with $x = 19$ is a very influential point—the other points alone give no indication of slope for the line. Seeing how widely scattered the y coordinates of the other points are, we cannot place too much faith in the y coordinate of the influential point; thus, we cannot depend on the slope of the line, so we cannot depend on the estimate when $x = 10$. (We also have no evidence as to whether or not a line is an appropriate model for this relationship.)

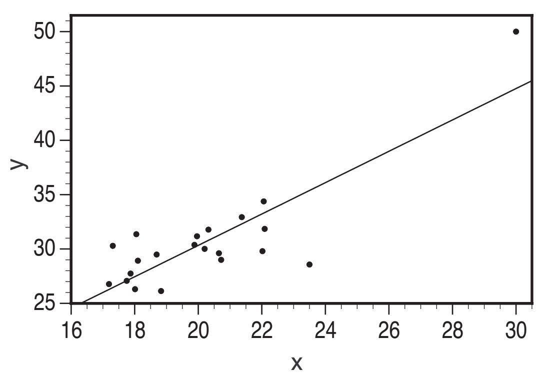


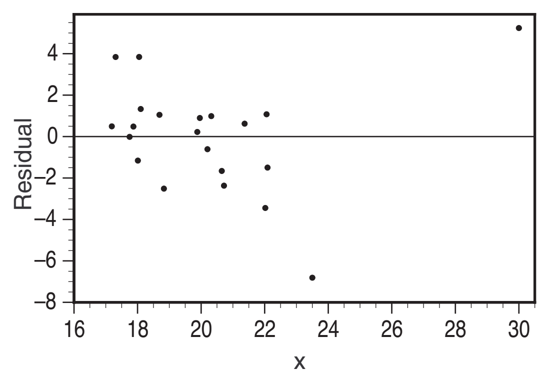
2.74. **(a)** The scatterplot (below, left) suggests a moderate positive linear relationship. **(b)** The regression equation is $\hat{y} = 17.38 + 0.6233x$. **(c)** The residual plot is below, right. **(d)** The regression explains $r^2 \doteq 27.4\%$ of the variation in y. **(e)** Student summaries will vary.



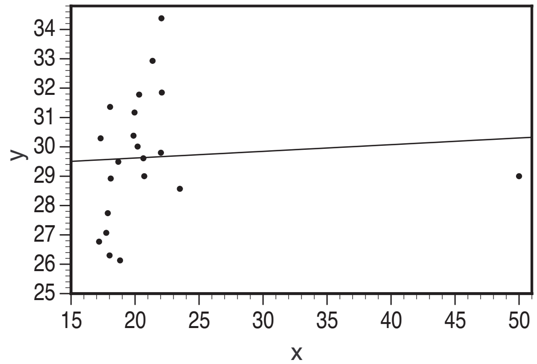


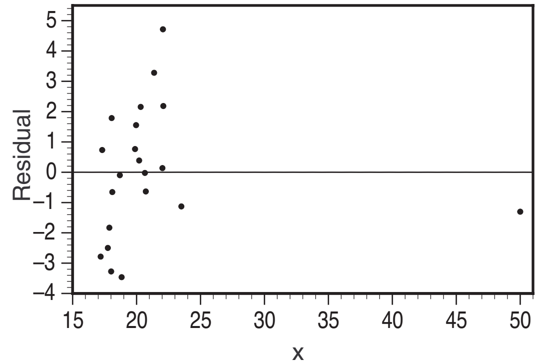
2.75. **(a)** The scatterplot (below, left) suggests a fairly strong positive linear relationship. **(b)** The regression equation is $\hat{y} = 1.470 + 1.4431x$. **(c)** The residual plot is below, right. The new point's residual is positive; the other residuals decrease as x increases. **(d)** The regression explains $r^2 \doteq 71.1\%$ of the variation in y. **(e)** The new point makes the relationship stronger, but its location has a large impact on the regression equation—both the slope and intercept changed substantially.





2.76. **(a)** The scatterplot (following page, left) gives little indication of a relationship between x and y. The regression equation is $\hat{y} = 29.163 + 0.02278x$; it explains only $r^2 \doteq 0.5\%$ of the variation in y. The residual plot (following page, right) tells a similar story to the first scatterplot—little evidence of a relationship. This new point does not fall along the same line as the other points, so it drastically weakens the relationship. **(b)** A point that does not follow the same pattern as the others can drastically change an association, and in extreme cases, can essentially make it disappear.

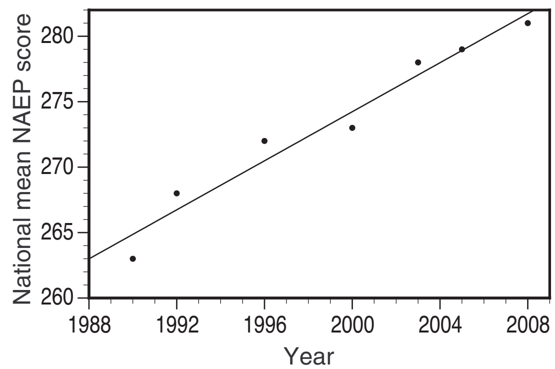




2.77. **(a)** When $x = 5$, $y = 12 + 6 \times 5 = 42$. **(b)** y increases by 6. (The change in y corresponding to a unit increase in x is the slope of this line.) **(c)** The intercept of this equation is 12.

2.78. **(a)** Time plot shown on the right, along with the regression line. **(b)** The means and standard deviations are $\bar{x} \doteq 1999.14$, $\bar{y} \doteq 273.43$, $s_x \doteq 6.7436$, and $s_y \doteq 6.4513$. With the correlation $r \doteq 0.9791$, the slope and intercept are

$$b_1 = r\, s_y/s_x \doteq 0.9366 \quad \text{and}$$
$$b_0 = \bar{y} - b_1\bar{x} \doteq -1599$$



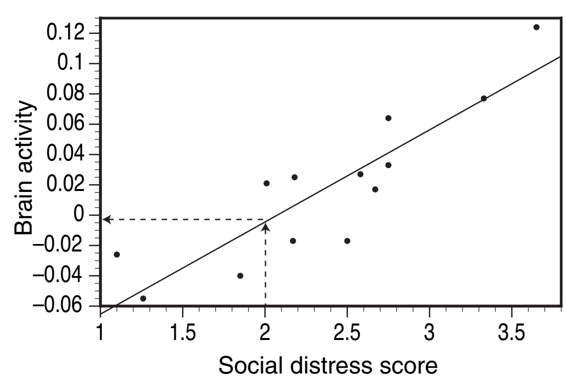
The equation is therefore $\hat{y} = -1599 + 0.9366x$; this line explains about $r^2 \doteq 95.9\%$ of the variation in score. **(c)** Obviously, the software regression line should be the same.

Note: *In examining student time plots, make sure they have a consistent scale on the horizontal axis; a common mistake is to ignore the variation in the gaps between years and leave an equal amount of horizontal space between each data point. Also, students may need to be reminded to carry extra digits in their computations of the slope and intercept.*

2.79. See also the solution to Exercise 2.33. **(a)** The regression equation is

$$\hat{y} = 0.06078x - 0.1261$$

(b) Based on the "up-and-over" method, most students will probably estimate that $\hat{y} \doteq 0$; the regression formula gives $\hat{y} = -0.0045$. **(c)** The correlation is $r \doteq 0.8782$, so the line explains $r^2 \doteq 77\%$ of the variation in brain activity.



2.80. The regression equations are $\hat{y} = -2.39 + 0.158x$ (Run 8903, 11.9 mg/s) and $\hat{y} = -1.45 + 0.0911x$ (Run 8905, 29.6 mg/s). Therefore, the growth rates are (respectively) 0.158 cm/minute and 0.0911 cm/minute; this suggests that the faster the water flows, the more slowly the icicles grow.

2.81. The means and standard deviations are $\bar{x} = 95$ min, $\bar{y} \doteq 12.6611$ cm, $s_x \doteq 53.3854$ min, and $s_y \doteq 8.4967$ cm; the correlation is $r \doteq 0.9958$.

For predicting length from time, the slope and intercept are $b_1 = r\,s_y/s_x \doteq 0.158$ cm/min and $a_1 = \bar{y} - b_1\bar{x} \doteq -2.39$ cm, giving the equation $\hat{y} = -2.39 + 0.158x$ (as in Exercise 2.80).

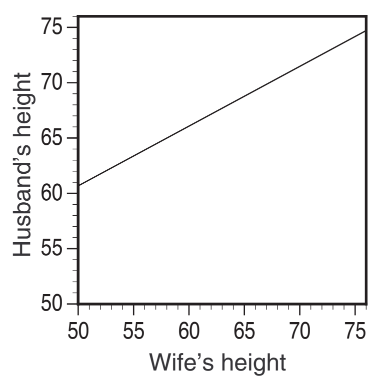
For predicting time from length, the slope and intercept are $b_2 = r\,s_x/s_y \doteq 6.26$ min/cm and $a_2 = \bar{x} - b_2\bar{y} \doteq 15.79$ min, giving the equation $\hat{x} = 15.79 + 6.26y$.

2.82. The means and standard deviations are: For lean body mass, $\bar{m} = 46.74$ and $s_m = 8.28$ kg, and for metabolic rate, $\bar{r} = 1369.5$ and $s_r = 257.5$ cal/day. The correlation is $r = 0.8647$. For predicting metabolic rate from body mass, the slope is $b_1 = r \cdot s_r/s_m \doteq 26.9$ cal/day per kg. For predicting body mass from metabolic rate, the slope is $b_2 = r \cdot s_m/s_r \doteq 0.0278$ kg per cal/day.

2.83. The correlation of IQ with GPA is $r_1 \doteq 0.634$; for self-concept and GPA, $r_2 \doteq 0.542$. IQ does a slightly better job; it explains about $r_1^2 \doteq 40.2\%$ of the variation in GPA, while self-concept explains about $r_2^2 \doteq 29.4\%$ of the variation.

2.84. Women's heights are the x-values; men's are the y-values. The slope is $b_1 = (0.5)(2.7)/2.5 = 0.54$ and the intercept is $b_0 = 68.5 - (0.54)(64.5) = 33.67$.

The regression equation is $\hat{y} = 33.67 + 0.54x$. Ideally, the scales should be the same on both axes. For a 67-inch tall wife, we predict the husband's height will be about 69.85 inches.

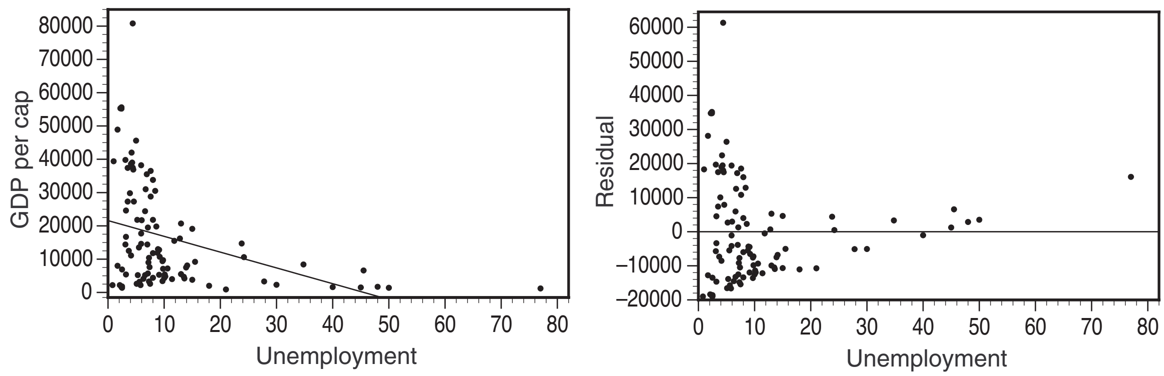


2.85. We have slope $b_1 = r\,s_y/s_x$ and intercept $b_0 = \bar{y} - b_1\bar{x}$, and $\hat{y} = b_0 + b_1x$, so when $x = \bar{x}$, $\hat{y} = b_0 + b_1\bar{x} = (\bar{y} - b_1\bar{x}) + b_1\bar{x} = \bar{y}$. (Note that the value of the slope does not actually matter.)

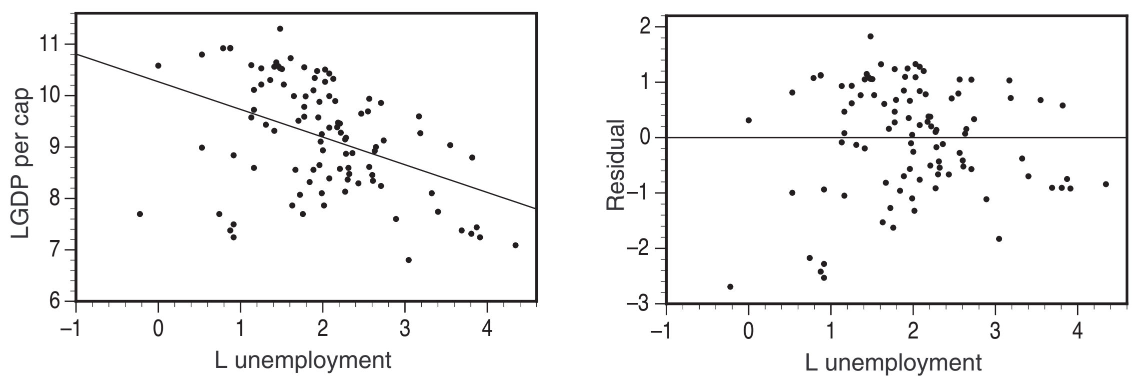
2.86. **(a)** $\bar{x} = 95$ min, $s_x \doteq 53.3854$ min, $\bar{y} \doteq 12.6611$ cm, and $s_y \doteq 8.4967$ cm. The correlation $r \doteq 0.9958$ has no units. **(b)** Multiply the old values of $\bar{y}$ and s_y by 2.54: $\bar{y} \doteq 32.1591$ and $s_y \doteq 21.5816$ inches. The correlation r is unchanged. **(c)** The slope is $r\,s_y/s_x$; with s_y from part (b), this gives $b_1 = 0.4025$ in/min. (Or multiply by 2.54 the appropriate slope from the solution to Exercise 2.80.)

2.87. $r = \sqrt{0.16} = 0.40$ (high attendance goes with high grades, so r must be positive).

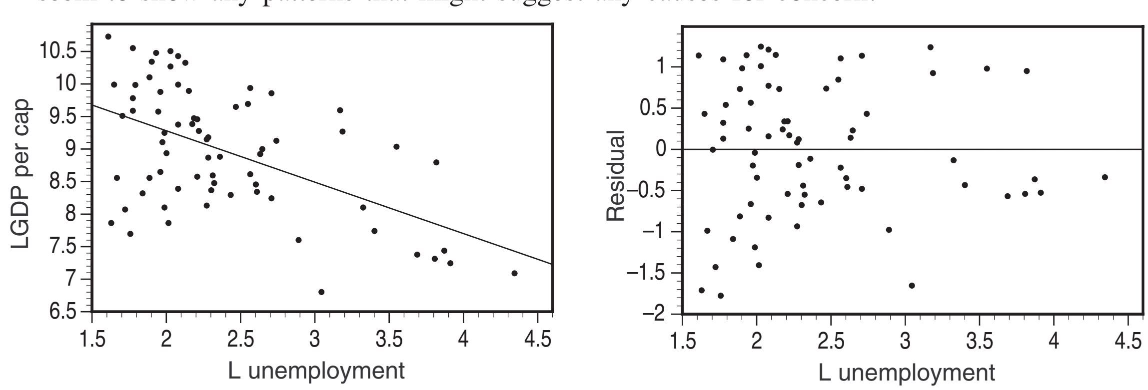
2.88. **(a)** The scatterplot (below, left) includes the regression line $\hat{y} = 21581 - 473.73x$. **(b)** The scatterplot does not suggest a linear association, so a regression line is not an appropriate summary of the relationship. **(c)** The residual plot (below, right) reveals—in a manner similar to the original scatterplot—that a line is not appropriate for this relationship. More specifically, the wide range of GDP-per-cap values for low unemployment rates suggest that there may be no useful relationship unless unemployment is sufficiently high.



2.89. **(a)** The scatterplot (below, left) includes the regression line $\hat{y} = 10.27 - 0.5382x$. **(b)** The scatterplot looks more linear than Figure 2.5, but a line may not be appropriate for all values of log unemployment. **(c)** In the residual plot (below, right), we see that there are more negative residuals on the left and right, with more positive residuals in the middle.

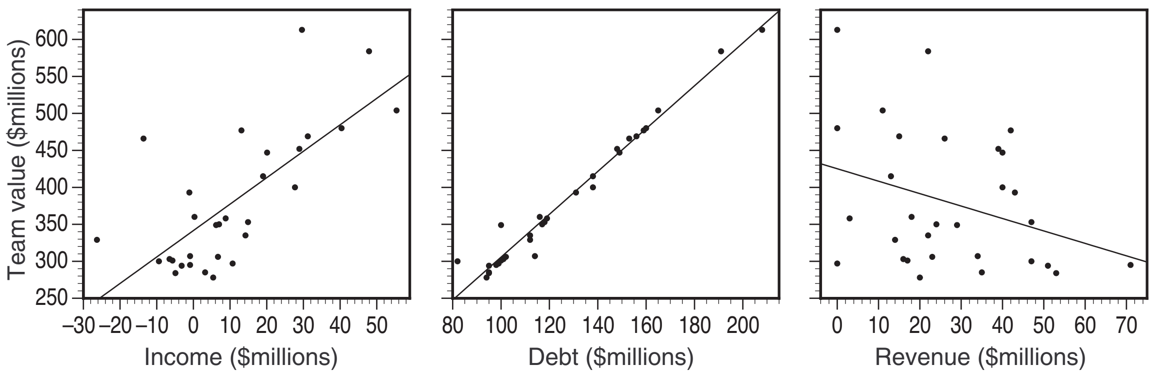


2.90. After removing countries with low unemployment rates, there are 70 countries left. **(a)** The scatterplot (below, left) includes the regression line $\hat{y} = 10.8614 - 0.7902x$. **(b)** A line seems appropriate for this set of countries, **(c)** The residual plot (below, right) does not seem to show any patterns that might suggest any causes for concern.



2.91. See the solutions to Exercise 2.34 (scatterplots) and Exercise 2.51 (correlations). Shown below are the three scatterplots, with regression lines; the equations for those lines are given on the right. The best regression line is clearly the one based on debt.

Explanatory variable	Equation	r^2
Income	$341.54 + 3.5760x$	51.5%
Debt	$16.02 + 2.8960x$	97.2%
Revenue	$425.17 - 1.6822x$	10.4%



2.92. For an NEA increase of 143 calories, the predicted fat gain is $\hat{y} = 3.505 - 0.00344 \times 143 \doteq 3.013$ kg, so the residual is $y - \hat{y} \doteq 3.2 - 3.013 = 0.187$ kg. This residual is positive because the actual fat gain was greater than the prediction.

2.93. The sum of the residuals is 0.01.

2.94. **(a)** It is impossible for all the residuals to be positive; some must be negative, because they will always sum to 0. **(b)** The direction of the relationship (positive or negative) has no connection to whether or not it is due to causation. **(c)** Lurking variables can be any kind of variables (and may be more likely to be explanatory rather than response).

2.95. **(a)** A high correlation means strong association, not causation. **(b)** Outliers in the y direction (and some other data points) will have large residuals. **(c)** It is not extrapolation if $1 \le x \le 5$.

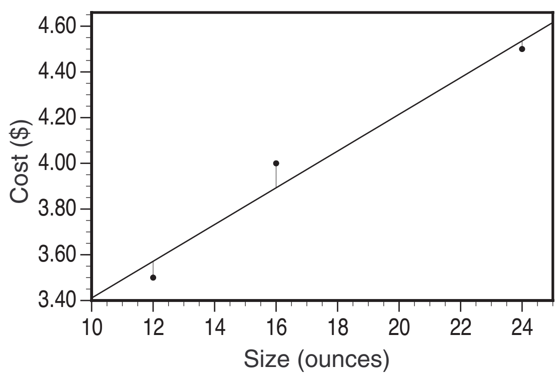
2.96. Both variables are indicators of the level of technological and economic development in a country; typically, they will both be low in a poorer, underdeveloped country, and they will both be high in a more affluent country.

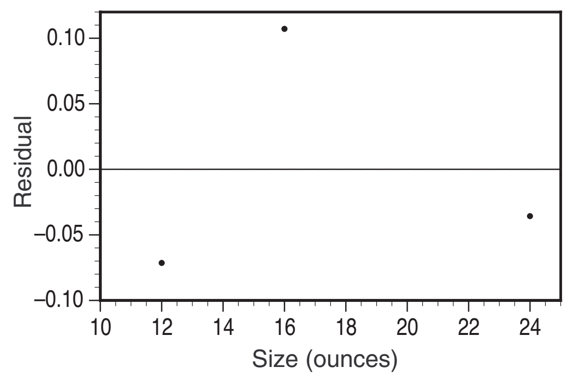
2.97. A reasonable explanation is that the cause-and-effect relationship goes in the other direction: Doing well makes students or workers feel good about themselves, rather than vice versa.

2.98. Patients suffering from more serious illnesses are more likely to go to larger hospitals (which may have more or better facilities) for treatment. They are also likely to require more time to recuperate afterwards.

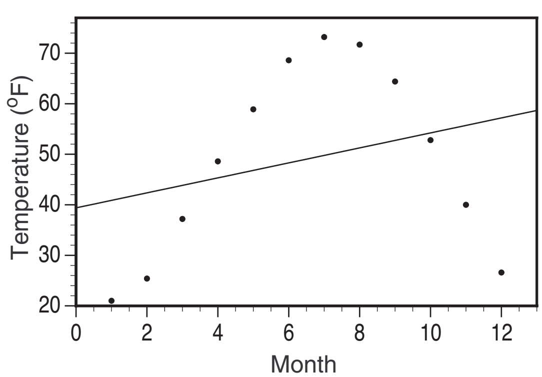
2.99. The explanatory and response variables were "consumption of herbal tea" and "cheerfulness/health." The most important lurking variable is social interaction; many of the nursing home residents may have been lonely before the students started visiting.

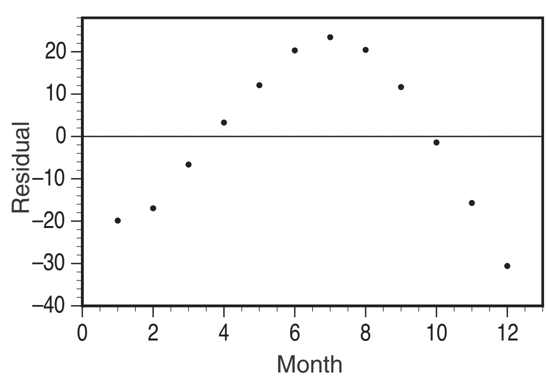
2.100. See also the solutions to Exercises 2.3 and 2.9. **(a)** Size should be on the horizontal axis because it is the explanatory variable. **(b)** The regression line is $\hat{y} = 2.6071 + 0.08036x$. **(c)** See the plot (next page, left). **(d)** Rounded to four decimal places, the residuals (as computed by software) are -0.0714, 0.1071, and -0.0357. It turns out that these three residuals add up to 0, no matter how much they are rounded. However, if they are computed by hand, and the slope and intercept in the regression equation have been rounded, there might be some roundoff error. **(e)** The middle residual is positive and the other two are negative, meaning that the 16-ounce drink costs more than the predicted value and the other two sizes cost less than predicted. Note that the residuals show the same pattern (relative to a horizontal line at 0) as the original points around the regression line.



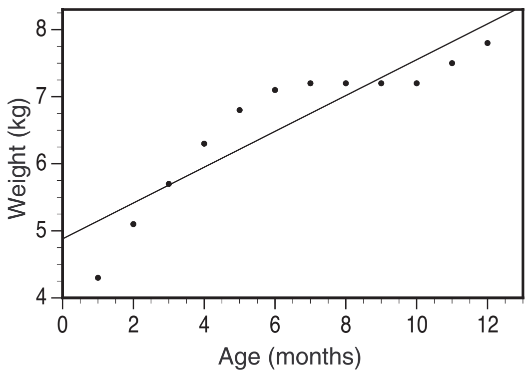


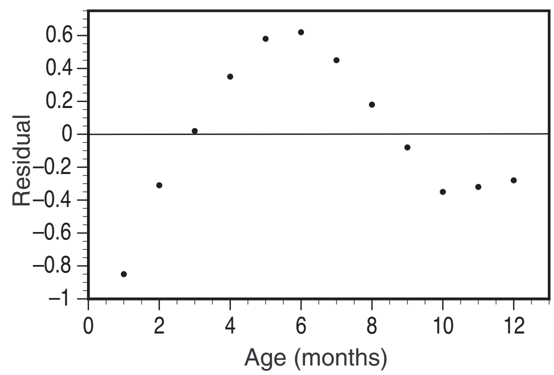
2.101. **(a)** The plot (below, left) is curved (low at the beginning and end of the year, high in the middle). **(b)** The regression line is $\hat{y} = 39.392 + 1.4832x$. It does not fit well because a line is poor summary of this relationship. **(c)** Residuals are negative for January through March and October through December (when actual temperature is less than predicted temperature), and positive from April to September (when it is warmer than predicted). **(d)** A similar pattern would be expected in any city that is subject to seasonal temperature variation. **(e)** Seasons in the Southern Hemisphere are reversed, so temperature would be cooler in the middle of the year.





2.102. **(a)** Below, left. **(b)** This line is not a good summary of the pattern; the scatterplot is curved rather than linear. **(c)** The sum is 0.01. The first two and last four residuals are negative, and those in the middle are positive. Plot below, right.





2.103. With individual children, the correlation would be smaller (closer to 0) because the additional variation of data from individuals would increase the "scatter" on the scatterplot, thus decreasing the strength of the relationship.

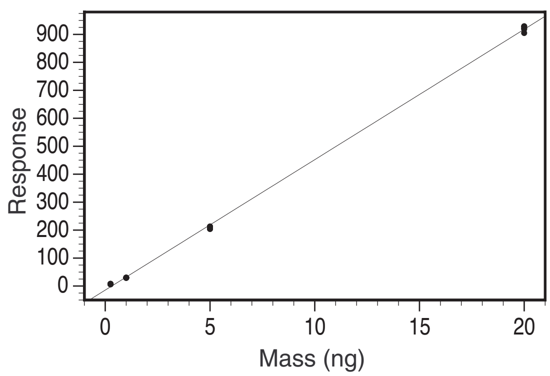
2.104. Presumably, those applicants who were hired would generally have been those who scored well on the test. As a result, we have little or no information on the job performance of those who scored poorly (and were therefore not hired). Those with higher test scores (who were hired) will likely have a range of performance ratings, so we will only see the various ratings for those with high scores, which will almost certainly show a weaker relationship than if we had performance ratings for all applicants.

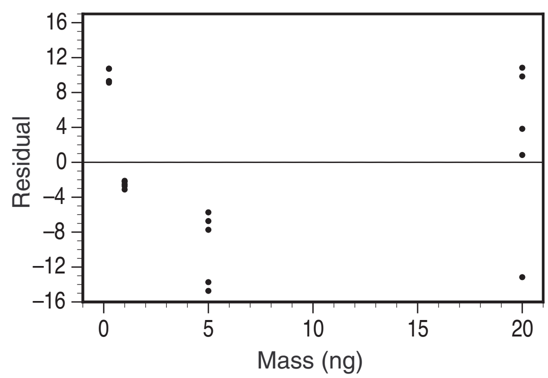
2.105. For example, a student who in the past might have received a grade of B (and a lower SAT score) now receives an A (but has a lower SAT score than an A student in the past). While this is a bit of an oversimplification, this means that today's A students are yesterday's A and B students, today's B students are yesterday's C students, and so on. Because of the grade inflation, we are not comparing students with equal abilities in the past and today.

2.106. A simple example illustrates this nicely: Suppose that everyone's current salary is their age (in thousands of dollars); for example, a 52-year-old worker makes \$52,000 per year. Everyone receives a \$500 raise each year. That means that in two years, every worker's income has increased by \$1000, but their age has increased by 2, so each worker's salary is now their age minus 1 (thousand dollars).

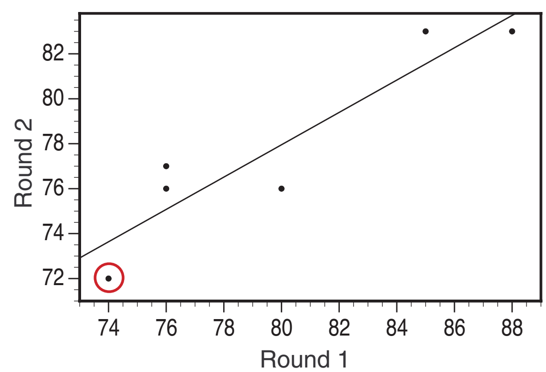
2.107. The correlation between BMR and fat gain is $r \doteq 0.08795$; the slope of the regression line is $b = 0.000811$ kg/cal. These both show that BMR is less useful for predicting fat gain. The small correlation suggests a very weak linear relationship (explaining less than 1% of the variation in fat gain). The small slope means that changes in BMR have very little impact on fat gain; for example, increasing BMR by 100 calories changes fat gain by only 0.08 kg.

2.108. **(a)** The scatterplot of the data is below on the left. (It is difficult to tell that there are 20 data points, because many of the points overlap.) **(b)** The regression equation is $\hat{y} = -14.4 + 46.6x$. **(c)** Residual plot below, right. The residuals for the extreme x-values ($x = 0.25$ and $x = 20.0$) are almost all positive; all those for the middle two x-values are negative.

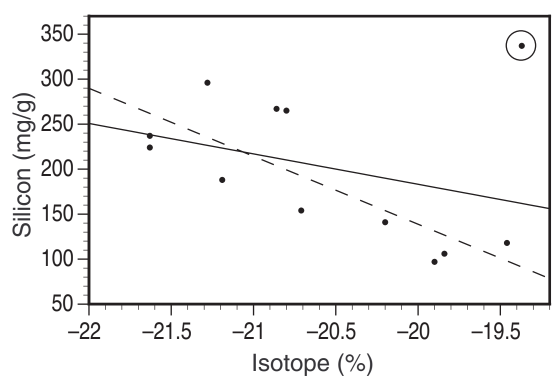




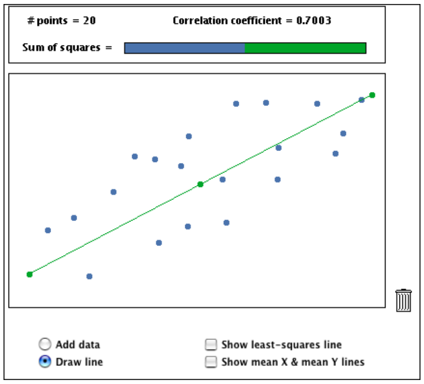
2.109. **(a)** Scatterplot on the right. **(b)** The plot shows a strong positive linear relationship. **(c)** The regression equation is $\hat{y} = 20.40 + 0.7194x$. **(d)** Hernandez's point is in the lower left—a logical place for the eventual champion.

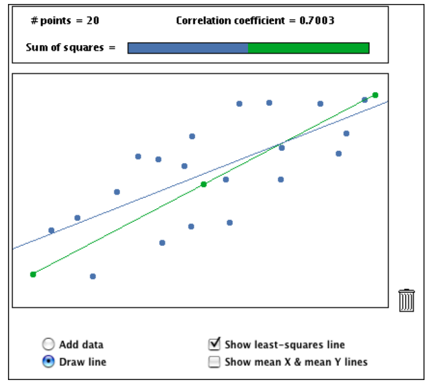


2.110. **(a)** Apart from the outlier—circled for part (b)—the scatterplot shows a moderate linear negative association. **(b)** With the outlier, $r = -0.3387$; without it, $r^* = -0.7866$. **(c)** The two regression formulas are $\hat{y} = -492.6 - 33.79x$ (the solid line, with all points) and $\hat{y} = -1371.6 - 75.52x$ (the dashed line, with the outlier omitted). The omitted point is also influential, as it has a noticeable impact on the line.

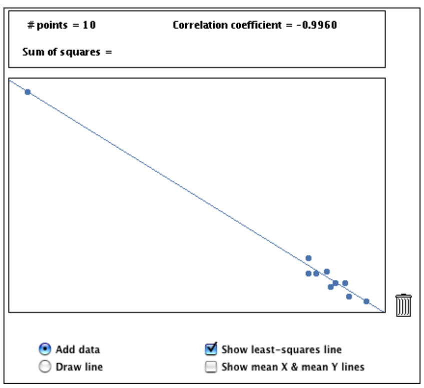


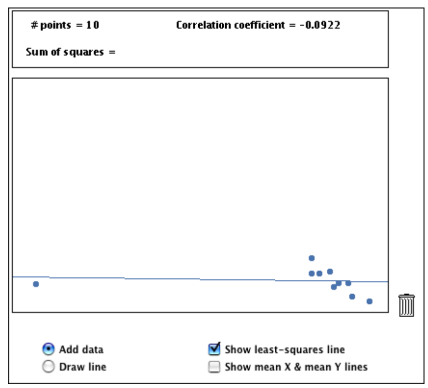
2.111. **(a)** Drawing the "best line" by eye is a very inaccurate process; few people choose the best line (although you can get better at it with practice). **(b)** Most people tend to overestimate the slope for a scatterplot with $r \doteq 0.7$; that is, most students will find that the least-squares line is less steep than the one they draw.



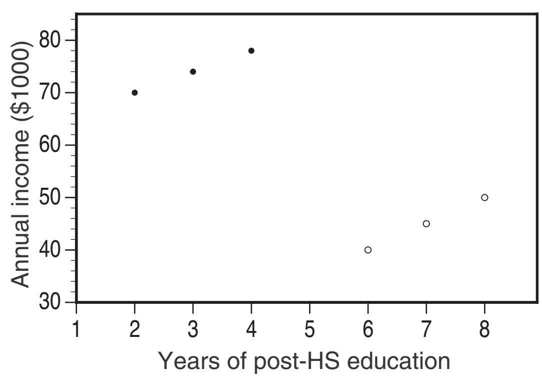


2.112. (a) Any point that falls exactly on the regression line will not increase the sum of squared vertical distances (which the regression line minimizes). Any other line—even if it passes through this new point—will necessarily have a higher total sum of squares. Thus, the regression line does not change. Possible output below, left. **(b)** Influential points are those whose x coordinates are outliers; this point is on the right side, while all others are on the left. Possible output below, right.



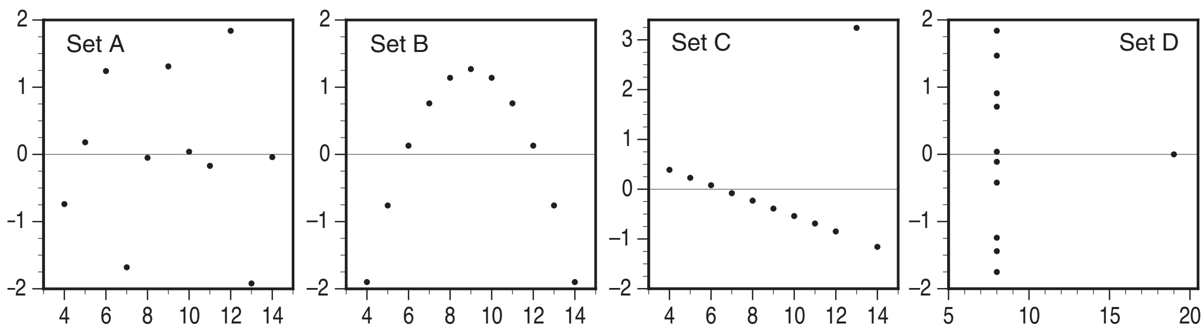


2.113. The plot shown is a very simplified (and not very realistic) example. Filled circles are economists in business; open circles are teaching economists. The plot should show positive association when either set of circles is viewed separately and should show a large number of bachelor's degree economists in business and graduate degree economists in academia.



2.114. See also the solution to Exercise 2.73. **(a)** Fits and residuals are listed below. (Students should find these using software.) **(b)** Plots below. **(c)** The residual plots confirm the observations made in Exercise 2.73: Regression is only appropriate for Set A.

Set A				Set B				Set C				Set D			
x	y	Fit	Resid.	x	y	Fit	Resid.	x	y	Fit	Resid.	x	y	Fit	Resid.
10	8.04	8.00	0.04	10	9.14	8.00	1.14	10	7.46	8.00	−0.54	8	6.58	7.00	−0.42
8	6.95	7.00	−0.05	8	8.14	7.00	1.14	8	6.77	7.00	−0.23	8	5.76	7.00	−1.24
13	7.58	9.50	−1.92	13	8.74	9.50	−0.76	13	12.74	9.50	3.24	8	7.71	7.00	0.71
9	8.81	7.50	1.31	9	8.77	7.50	1.27	9	7.11	7.50	−0.39	8	8.84	7.00	1.84
11	8.33	8.50	−0.17	11	9.26	8.50	0.76	11	7.81	8.50	−0.69	8	8.47	7.00	1.47
14	9.96	10.00	−0.04	14	8.10	10.00	−1.90	14	8.84	10.00	−1.16	8	7.04	7.00	0.04
6	7.24	6.00	1.24	6	6.13	6.00	0.13	6	6.08	6.00	0.08	8	5.25	7.00	−1.75
4	4.26	5.00	−0.74	4	3.10	5.00	−1.90	4	5.39	5.00	0.39	8	5.56	7.00	−1.44
12	10.84	9.00	1.84	12	9.13	9.00	0.13	12	8.15	9.00	−0.85	8	7.91	7.00	0.91
7	4.82	6.50	−1.68	7	7.26	6.50	0.76	7	6.42	6.50	−0.08	8	6.89	7.00	−0.11
5	5.68	5.50	0.18	5	4.74	5.50	−0.76	5	5.73	5.50	0.23	19	12.50	12.50	0.00



2.115. There are 1684 female binge drinkers in the table; 8232 female students are not binge drinkers.

2.116. There are $1684 + 8232 = 9916$ women in the study. The number of students who are not binge drinkers is $5550 + 8232 = 13{,}782$.

2.117. Divide the number of non-bingeing females by the total number of students:

$$\frac{8232}{17{,}096} \doteq 0.482$$

2.118. Use the numbers in the right-hand column of the table in Example 2.28. Divide the counts of bingeing and non-bingeing students by the total number of students:

$$\frac{3314}{17{,}096} \doteq 0.194 \quad \text{and} \quad \frac{13{,}782}{17{,}096} \doteq 0.806$$

2.119. This is a conditional distribution; take the number of bingeing males divided by the total number of males: $\frac{1630}{7180} \doteq 0.227$.

2.120. The first computation was performed in the previous solution; for the second, take the number of non-bingeing males divided by the total number of males: $\frac{5550}{7180} \doteq 0.773$.

2.121. **(a)** There are $151 + 148 = 299$ "high exercisers," of which $\frac{151}{299} \doteq 50.5\%$ get enough sleep and 49.5% (the rest) do not. **(b)** There are $115 + 242 = 357$ "low exercisers," of which $\frac{115}{357} \doteq 32.2\%$ get enough sleep and 67.8% (the rest) do not. **(c)** Those who exercise more than the median are more likely to get enough sleep.

Note: *This question is asking for the conditional distribution of sleep within each exercise group. The next question asks for the conditional distribution of exercise within each sleep group.*

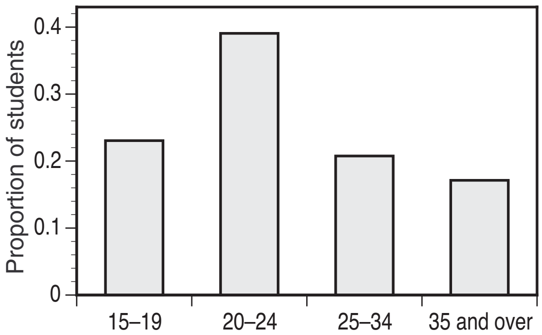
2.122. **(a)** There are $151 + 115 = 266$ students who get enough sleep, of which $\frac{151}{266} \doteq 56.8\%$ are high exercisers and 43.2% (the rest) are low exercisers. **(b)** There are $148 + 242 = 390$ students who do not get enough sleep, of which $\frac{148}{390} \doteq 37.9\%$ are high exercisers and 62.1% (the rest) are low exercisers. **(c)** Students who get enough sleep are more likely to be high exercisers. **(d)** Preferences will vary. In particular, note that one can make the case for a cause-and-effect relationship in either direction between these variables.

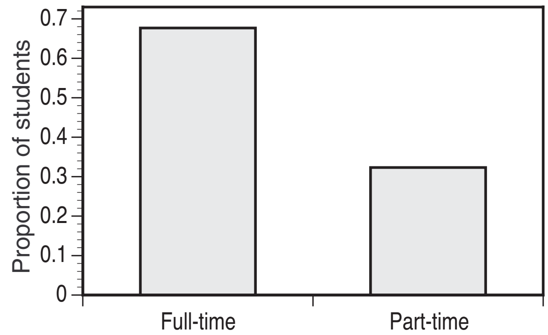
2.123. $\frac{63}{2100} = 3.0\%$ of Hospital A's patients died, compared with $\frac{16}{800} = 2.0\%$ at Hospital B.

2.124. **(a)** For patients in poor condition, Hospital A lost $\frac{57}{1500} = 3.8\%$, while Hospital B lost $\frac{8}{200} = 4.0\%$. Hospital A did slightly better with patients in poor condition. **(b)** For patients in good condition, Hospital A lost $\frac{6}{600} = 1.0\%$ and Hospital B lost $\frac{8}{600} \doteq 1.33\%$. Again, Hospital A did slightly better. **(c)** Hospital A appears to be the safer choice. **(d)** More than 70% of Hospital A's patients arrive in poor condition, compared to 25% at Hospital B, so A's survival rate is lower overall because these patients are more likely to die in the hospital.

2.125. Bar graphs are on the following page. **(a)** There are about 3,388,000 full-time college students aged 15 to 19. (Note that numbers are in thousands.) **(b)** The joint distribution is found by dividing each number in the table by 16,388 (the total of all the numbers). These proportions are given in italics in the table on the right. For example, $\frac{3388}{16,388} \doteq 0.2067$, meaning that about 20.7% of all college students are full-time and aged 15 to 19. **(c)** The marginal distribution of age is found by dividing the *row* totals by 16,388; they are in the right margin of the table and the graph on the left following. For example, $\frac{3777}{16,388} \doteq 0.2305$, meaning that about 23% of all college students are aged 15 to 19. **(d)** The marginal distribution of status is found by dividing the *column* totals by 16,388; they are in the bottom margin of the table and the graph on the right following. For example, $\frac{11,091}{16,388} \doteq 0.6768$, meaning that about 67.7% of all college students are full-time.

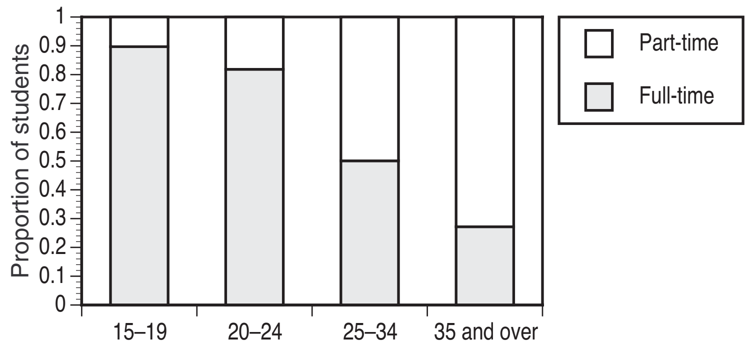
	FT	PT	
15–19	3388 *0.2067*	389 *0.0237*	3777 0.2305
20–24	5238 *0.3196*	1164 *0.0710*	6402 0.3907
25–34	1703 *0.1039*	1699 *0.1037*	3402 0.2076
35+	762 *0.0465*	2045 *0.1248*	2807 0.1713
	11091 0.6768	5297 0.3232	16388





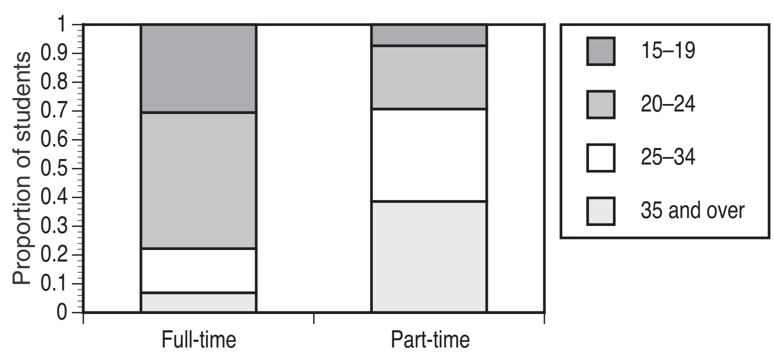
2.126. Refer to the counts in the solution to Exercise 2.125. For each age category, the conditional distribution of status is found by dividing the counts in that row by that row total. For example, $\frac{3388}{3777} \doteq 0.8970$ and $\frac{389}{3777} \doteq 0.1030$, meaning that of all college students in the 15–19 age range, about 89.7% are full-time, and the rest (10.3%) are part-time. Note that each pair of numbers should add up to 1 (except for rounding error, but with only two numbers, that rarely happens). The complete table is shown below, along with one possible graphical presentation. We see that the older the students are, the more likely they are to be part-time.

	FT	PT
15–19	0.8970	0.1030
20–24	0.8182	0.1818
25–34	0.5006	0.4994
35+	0.2715	0.7285



2.127. Refer to the counts in the solution to Exercise 2.125. For each status category, the conditional distribution of age is found by dividing the counts in that column by that column total. For example, $\frac{3388}{11,091} \doteq 0.3055$, $\frac{5238}{11,091} \doteq 0.4723$, etc., meaning that of all full-time college students, about 30.55% are aged 15 to 19, 47.23% are 20 to 24, and so on. Note that each set of four numbers should add up to 1 (except for rounding error). Graphical presentations may vary; one possibility is shown below. We see that full-time students are dominated by younger ages, while part-time students are more likely to be older. (This is essentially the same observation made in the previous exercise, seen from a different viewpoint.)

	FT	PT
15–19	0.3055	0.0734
20–24	0.4723	0.2197
25–34	0.1535	0.3207
35+	0.0687	0.3861



2.128. Two examples are shown on the right. In general, choose a to be any number from 0 to 200, and then all the other entries can be determined.

50	150
150	50

175	25
25	175

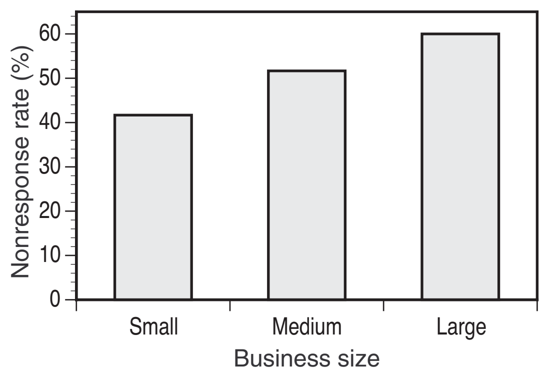
Note: *This is why we say that such a table has "one degree of freedom": We can make one (nearly) arbitrary choice for the first number, and then have no more decisions to make.*

2.129. To construct such a table, we can start by choosing values for the row and column sums $r1, r2, r3, c1, c2, c3$, as well as the grand total N. Note that $N = r1 + r2 + r3 = c1 + c2 + c3$, so we only have five choices to make. Then, find each count $a, b, c, d, e, f, g, h, i$ by taking the corresponding *row* total, times the corresponding *column* total, divided by the *grand* total. For example, $a = r1 \times c1/N$ and $f = r2 \times c3/N$. Of course, these counts should be whole numbers, so it may be necessary to make adjustments in the row and column totals to meet this requirement.

a	b	c	$r1$
d	e	f	$r2$
g	h	i	$r3$
$c1$	$c2$	$c3$	N

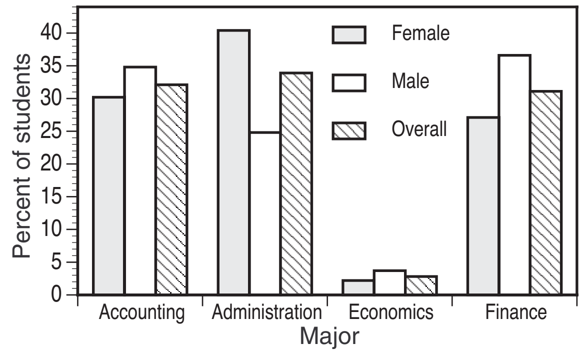
The simplest such table would have all nine counts $(a, b, c, d, e, f, g, h, i)$ equal to one another.

2.130. **(a)** Overall, $\frac{125+155+180}{900} \doteq 51.1\%$ did not respond. **(b)** Generally, the larger the business, the less likely it was to respond: $\frac{125}{300} \doteq 41.7\%$ of small businesses, $\frac{155}{300} \doteq 51.7\%$ of medium-sized businesses, and $\frac{180}{300} = 60.0\%$ of large businesses did not respond. **(c)** At right. **(d)** Of the 440 total responses, $\frac{175}{440} \doteq 39.8\%$ came from small businesses, $\frac{145}{440} \doteq 33.0\%$ from medium-sized businesses, and $\frac{120}{440} \doteq 27.3\%$ from large businesses. **(e)** No: Almost 40% of respondents were small businesses, while just over a quarter of all responses come from large businesses.

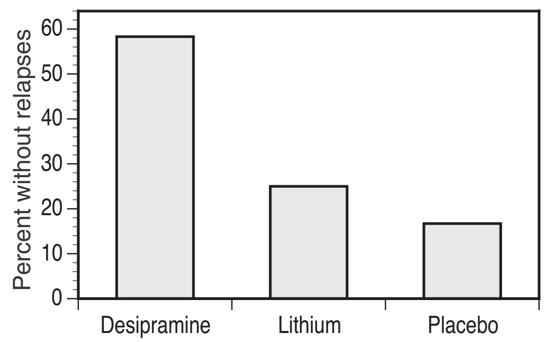


2.131. **(a)** Use column percents, e.g., $\frac{68}{225} \doteq 30.22\%$ of females are in administration. See table and graph below. The biggest difference between women and men is in Administration: A higher percentage of women chose this major. Meanwhile, a greater proportion of men chose other fields, especially Finance. **(b)** There were 386 responses; $\frac{336}{722} \doteq 46.5\%$ did not respond.

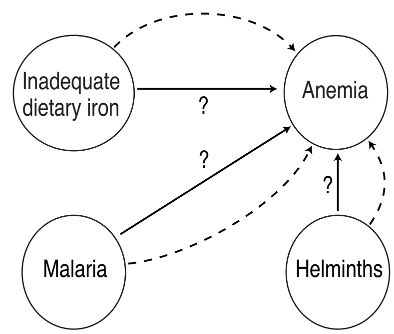
	Female	Male	Overall
Accting.	30.22%	34.78%	32.12%
Admin.	40.44%	24.84%	33.94%
Econ.	2.22%	3.70%	2.85%
Fin.	27.11%	36.65%	31.09%



2.132. $\frac{14}{24} \doteq 58.33\%$ of desipramine users did not have a relapse, while $\frac{6}{24} = 25\%$ of lithium users and $\frac{4}{24} \doteq 16.67\%$ of those who received placebos succeeded in breaking their addictions. Desipramine seems to be effective. Note that use of percentages is not as crucial here as in other cases because each drug was given to 24 addicts.

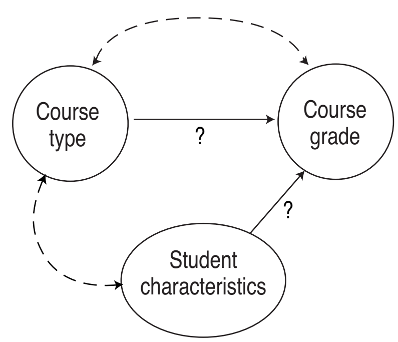


2.133. This is a case of confounding: The association between dietary iron and anemia is difficult to detect because malaria and helminths also affect iron levels in the body.

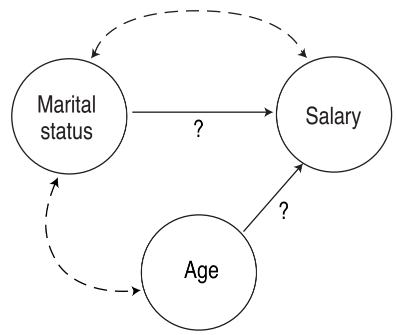


2.134. Opinions will vary; one can argue for causation both ways, and the truth is probably that both conditions exacerbate one another.

2.135. Responses will vary. For example, students who choose the online course might have more self-motivation or better computer skills. A diagram is shown on the right; the generic "Student characteristics" might be replaced with something more specific.

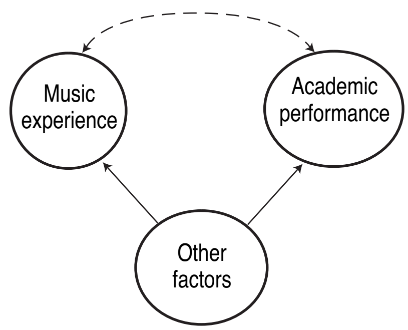


2.136. Age is one lurking variable: Married men would generally be older than single men, so they would have been in the work force longer and therefore had more time to advance in their careers. The diagram shown on the right shows this lurking variable; other variables could also be shown in place of "age."



2.137. No; self-confidence and improving fitness could be a common response to some other personality trait, or high self-confidence could make a person more likely to join the exercise program.

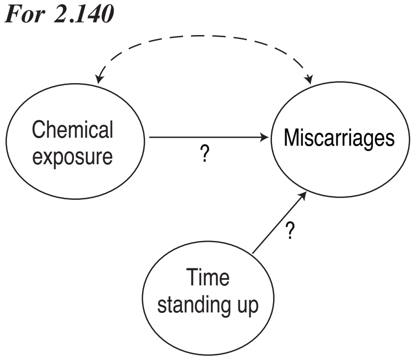
2.138. Students with music experience may have other advantages (wealthier parents, better school systems, and so forth). That is, experience with music may have been a "symptom" (common response) of some other factor that also tends to cause high grades.



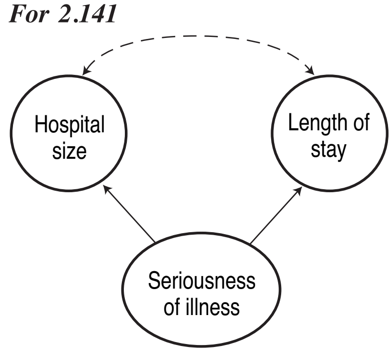
2.139. Two possibilities are that they might perform better simply because this is their second attempt or because they feel better prepared as a result of taking the course (whether or not they really *are* better prepared).

2.140. The diagram below illustrates the confounding between exposure to chemicals and standing up.

For 2.140



For 2.141



2.141. Patients suffering from more serious illnesses are more likely to go to larger hospitals (which may have more or better facilities) for treatment. They are also likely to require more time to recuperate afterwards.

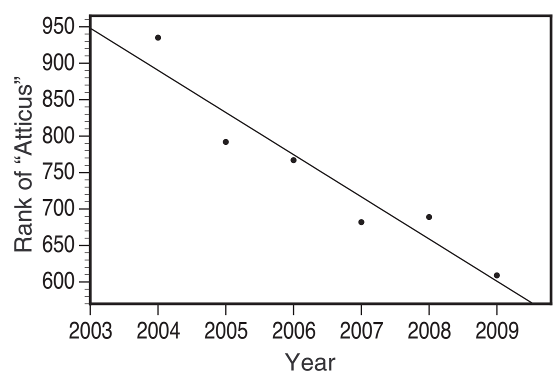
2.142. Spending more time watching TV means that *less* time is spent on other activities; this may suggest lurking variables. For example, perhaps the parents of heavy TV watchers do not spend as much time at home as other parents. Also, heavy TV watchers would typically not get as much exercise.

2.143. In this case, there may be a causative effect, but in the direction opposite to the one suggested: People who are overweight are more likely to be on diets and so choose artificial sweeteners over sugar. (Also, heavier people are at a higher risk to develop diabetes; if they do, they are likely to switch to artificial sweeteners.)

2.144. **(a)** Statements such as this typically mean that the risk of dying *at a given age* is half as great; that is, given two groups of the same age, where one group walks and the other does not, the walkers are half as likely to die in (say) the next year. **(b)** Men who choose to walk might also choose (or have chosen, earlier in life) other habits and behaviors that reduce mortality.

2.145. This is an observational study—students choose their "treatment" (to take or not take the refresher sessions).

2.146. **(a)** Time plot on the right. **(b)** The regression equation is $\hat{y} = 116779 - 57.83x$. **(c)** The plot shows a clear negative association, and the slope of the regression line says that the rank is decreasing at an average rate of about 58 per year. Because a *lower* rank means *higher* popularity, this means that "Atticus" is getting more popular.



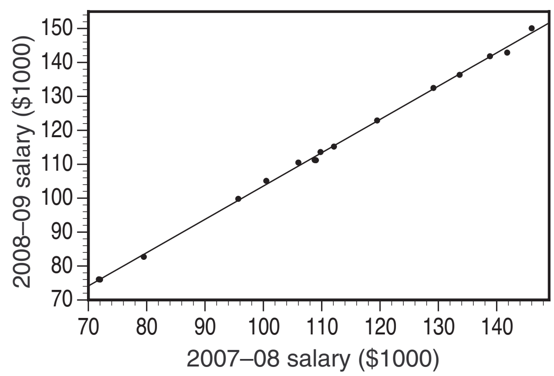
2.148. **(a)** The scatterplot shows a positive, curved relationship. **(b)** The regression explains about $r^2 \doteq 98.3\%$ of the variation in salary. While this indicates that the relationship is strong, and *close* to linear, we can see from that scatterplot that the actual relationship is curved.

2.149. **(a)** The residuals are positive at the beginning and end, and negative in the middle. **(b)** The behavior of the residuals agrees with the curved relationship seen in Figure 2.30.

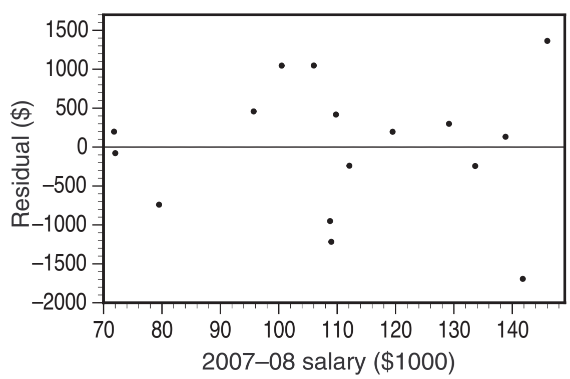
2.150. **(a)** Both plots show a positive association, but the log-salary plot is linear rather than curved. **(b)** While the residuals in Figure 2.31 were positive at the beginning and end, and negative in the middle, the log-salary residuals in Figure 2.33 show no particular pattern.

2.151. **(a)** The regression equation for predicting salary from year is $\hat{y} = 41.253 + 3.9331x$; for $x = 25$, the predicted salary is $\hat{y} \doteq 139.58$ thousand dollars, or about \$139,600. **(b)** The log-salary regression equation is $\hat{y} = 3.8675 + 0.04832x$. With $x = 25$, we have $\hat{y} \doteq 5.0754$, so the predicted salary is $e^{\hat{y}} \doteq 160.036$, or about \$160,040. **(c)** Although both predictions involve extrapolation, the second is more reliable, because it is based on a linear fit to a linear relationship. **(d)** Interpreting relationships without a plot is risky. **(e)** Student summaries will vary, but should include comments about the importance of looking at plots, and the risks of extrapolation.

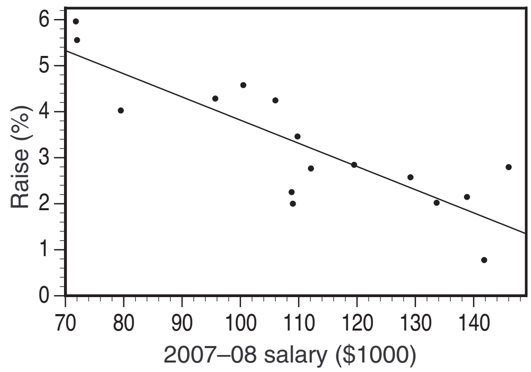
2.152. **(a)** The scatterplot on the right includes the regression line given in the next exercise. **(b)** The plot shows a very strong positive linear relationship. **(c)** The correlation between the two variables is $r \doteq 0.9993$, so regression explains $r^2 \doteq 99.9\%$ of the variation in 2008–09 salaries.

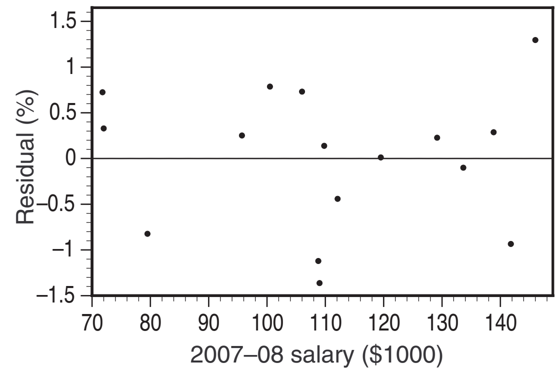


2.153. **(a)** The regression equation is $\hat{y} = 5403 + 0.9816x$. **(b)** The residual plot on the right reveals no causes for concern; in particular, there are no clear outliers or influential observations.



2.154. **(a)** A spreadsheet or other software is the best way to do these computations. As a check, the first two raises are $\frac{142{,}900-141{,}800}{141{,}800} \doteq 0.78\%$ and $\frac{113{,}600-109{,}800}{109{,}800} \doteq 3.46\%$. The scatterplot (following page, left) shows a moderately strong negative linear relationship. **(b)** The regression equation is $\hat{y} = 8.8536 - 0.00005038x$. **(c)** A plot of residuals versus 2007–08 salaries reveals no outliers or other causes for concern. **(d)** The data do show that, in general, those with lower salaries are given greater percentage raises. Students might observe, for example, that the regression explains 71.1% of the variation in raise. In addition, the regression slope tells us that on the average, an individual's raise decreases by about 0.5% for each additional \$10,000 earned in 2007–08, because $0.00005038 \times \$10{,}000 \doteq 0.5\%$. (Note that the slope is in units of percent per dollar.)

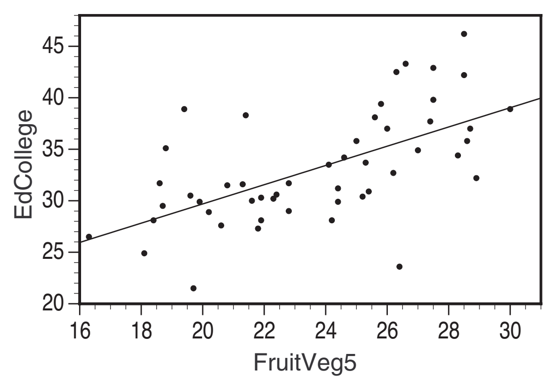




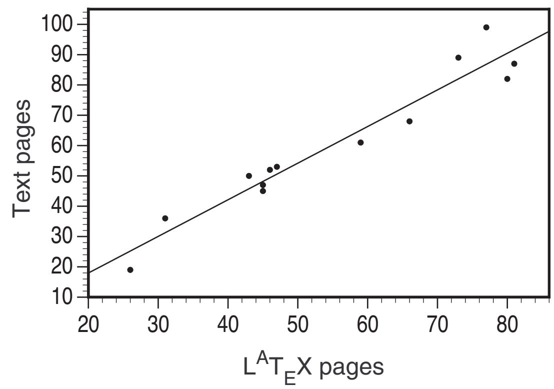
2.155. A school that accepts weaker students but graduates a higher-than-expected number of them would have a positive residual, while a school with a stronger incoming class but a lower-than-expected graduation rate would have a negative residual. It seems reasonable to measure school quality by how much benefit students receive from attending the school.

2.156. **(a)** The association is negative and roughly linear. This seems reasonable because a low number of smokers suggests that the state's population is health-conscious, so we might expect more people in that state to have healthy eating habits. **(b)** The correlation is $r \doteq -0.5503$. **(c)** Utah is the farthest point to the left (that is, it has the lowest smoking rate) and lies well below the line (i.e., the proportion of adults who eat fruits and vegetables is lower than we would expect). **(d)** California has the second-lowest smoking rate and one of the highest fruit/vegetable rates. This point lies above the line, meaning that the proportion of California adults who eat fruits and vegetables is higher than we would expect.

2.157. **(a)** The scatterplot shows a moderate positive association. **(b)** The regression line $\hat{y} = 11.00 + 0.9344x$ fits the overall trend. (However, there is some hint of a curve in the plot, because most of the points in the middle lie below the line.) **(c)** For example, a state whose point falls above the line has a higher percent of college graduates than we would expect based on the percent who eat 5 servings of fruits and vegetables. **(d)** No; association is not evidence of causation.



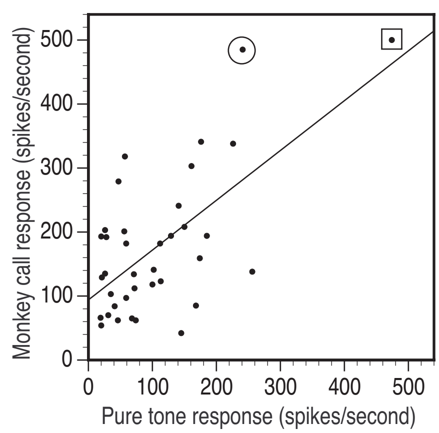
2.158. **(a)** The plot shows a fairly strong positive linear association. **(b)** The regression equation is $\hat{y} = -6.202 + 1.2081x$. **(c)** If $x = 62$ pages, we predict $\hat{y} \doteq 68.7$ pages.



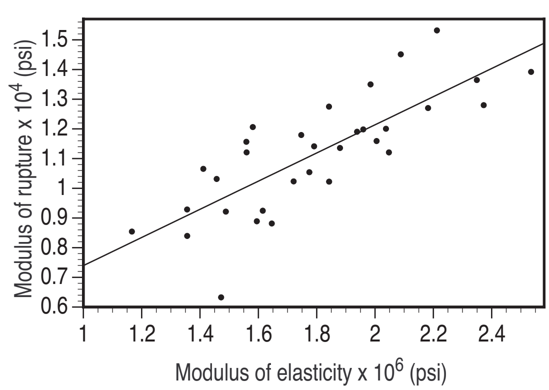
2.161. These results support the idea (the slope is negative, so variation decreases with increasing diversity), but the relationship is only moderately strong ($r^2 = 0.34$, so diversity only explains 34% of the variation in population variation).

Note: *That last parenthetical comment is awkward and perhaps confusing, but is consistent with similar statements interpreting* r^2.

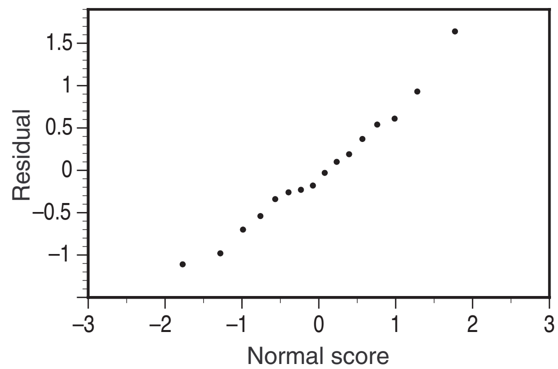
2.162. **(a)** One possible measure of the difference is the mean response: 106.2 spikes/second for pure tones and 176.6 spikes/second for monkey calls—an average of an additional 70.4 spikes/second. **(b)** The regression equation is $\hat{y} = 93.9 + 0.778x$. The third point (pure tone 241, call 485 spikes/second) has the largest residual; it is circled. The first point (474 and 500 spikes/second) is an outlier in the x direction; it is marked with a square. **(c)** The correlation drops only slightly (from 0.6386 to 0.6101) when the third point is removed; it drops more drastically (to 0.4793) without the first point. **(d)** Without the first point, the line is $\hat{y} = 101 + 0.693x$; without the third point, it is $\hat{y} = 98.4 + 0.679x$.



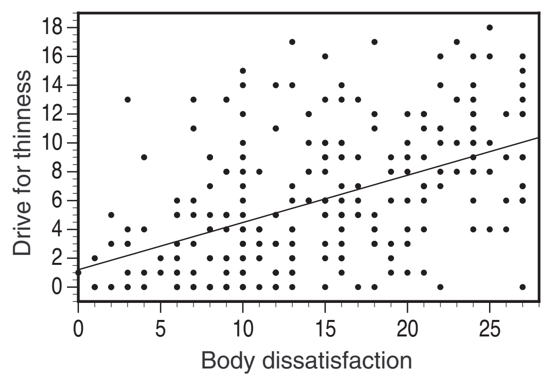
2.163. On the right is a scatterplot of MOR against MOE, showing a moderate linear positive association. The regression equation is $\hat{y} = 2653 + 0.004742x$; this regression explains $r^2 = 0.6217 \doteq 62\%$ of the variation in MOR. So, we can use MOE to get fairly good (though not perfect) predictions of MOR.



2.164. The quantile plot (right) is reasonably close to a straight line, so we have little reason to doubt that they come from a Normal distribution.

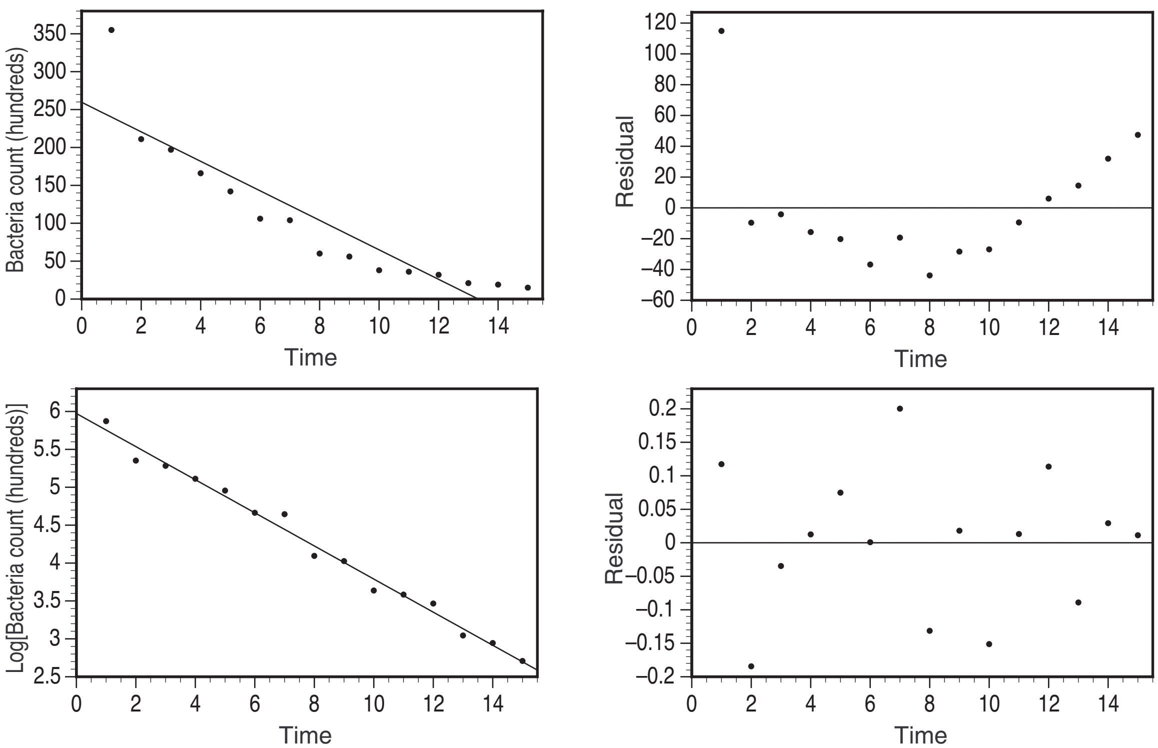


2.165. **(a)** The scatterplot is on the right. **(b)** The regression equation is $\hat{y} = 1.2027 + 0.3275x$. As we see from the scatterplot, the relationship is not too strong; the correlation ($r = 0.4916$, $r^2 = 0.2417$) confirms this.



2.166. **(a)** Yes: The two lines appear to fit the data well. There do not appear to be any outliers or influential points. **(b)** Compare the slopes: before—0.189; after—0.157. (The units for these slopes are 100 ft^3/day per degree-day/day; for students who are comfortable with units, 18.9 ft^3 vs. 15.7 ft^3 would be a better answer.) **(c)** Before: $\hat{y} = 1.089 + 0.189(35) = 7.704 = 770.4$ ft^3. After: $\hat{y} = 0.853 + 0.157(35) = 6.348 = 634.8$ ft^3. **(d)** This amounts to an additional $(\$1.20)(7.704 - 6.348) = \$1.63$ per day, or \$50.44 for the month.

2.167. **(a)** Shown below are plots of count against time, and residuals against time for the regression, which gives the formula $\hat{y} = 259.58 - 19.464x$. Both plots suggest a curved relationship rather than a linear one. **(b)** With natural logarithms, the regression equation is $\hat{y} = 5.9732 - 0.2184x$; with common logarithms, $\hat{y} = 2.5941 - 0.09486x$. The second pair of plots below show the (natural) logarithm of the counts against time, suggesting a fairly linear relationship, and the residuals against time, which shows no systematic pattern. (If common logarithms are used instead of natural logs, the plots will look the same, except the vertical scales will be different.) The correlations confirm the increased linearity of the log plot: $r^2 = 0.8234$ for the original data, $r^2 = 0.9884$ for the log-data.

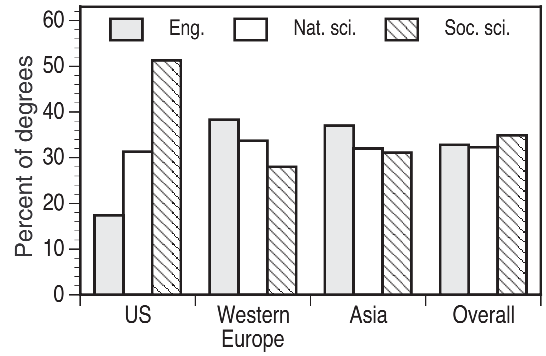


2.168. Note that $\bar{y} = 46.6 + 0.41\bar{x}$. We predict that Octavio will score 4.1 points above the mean on the final exam: $\hat{y} = 46.6 + 0.41(\bar{x} + 10) = 46.6 + 0.41\bar{x} + 4.1 = \bar{y} + 4.1$. (Alternatively, because the slope is 0.41, we can observe that an increase of 10 points on the midterm yields an increase of 4.1 on the predicted final-exam score.)

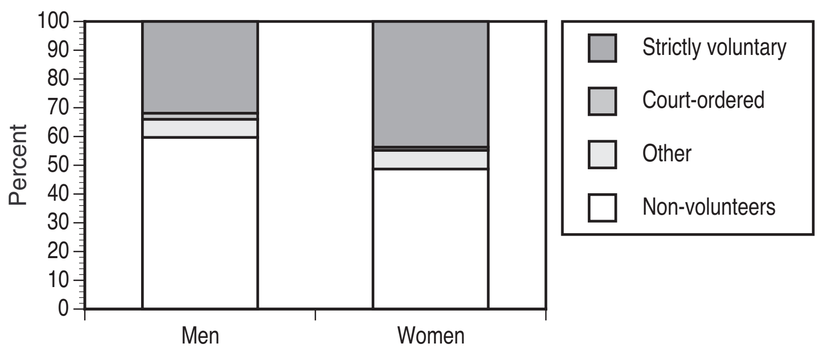
2.169. Number of firefighters and amount of damage both increase with the seriousness of the fire (i.e., they are common responses to the fire's seriousness).

2.170. Compute column percents; for example, $\frac{61,941}{355,265} \doteq 17.44\%$ of those U.S. degrees considered in this table are in engineering. See table and graph at right. We observe that there are considerably more social science degrees and fewer engineering degrees in the United States. The Western Europe and Asia distributions are similar.

Field	United States	Western Europe	Asia	Overall
Eng.	17.44%	38.26%	36.96%	32.78%
Nat. sci.	31.29%	33.73%	31.97%	32.29%
Soc. sci.	51.28%	28.01%	31.07%	34.93%

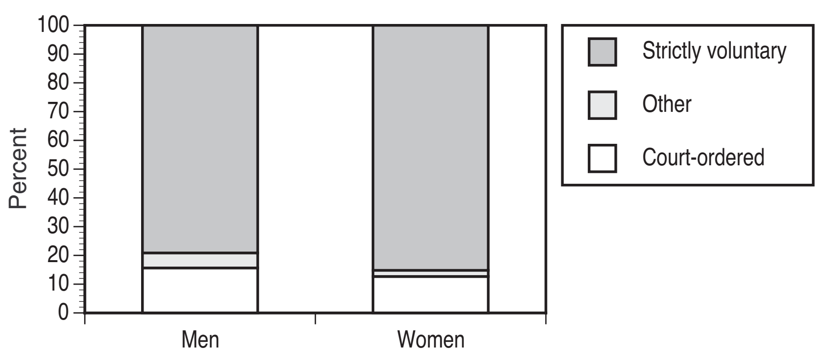


2.171. Different graphical presentations are possible; one is shown below. More women perform volunteer work; the notably higher percentage of women who are "strictly voluntary" participants accounts for the difference. (The "court-ordered" and "other" percentages are similar for men and women.)



2.172. Table shown on the right; for example, $\frac{31.9\%}{40.3\%} \doteq 79.16\%$. The percents in each row sum to 100%, with no rounding error for up to four places after the decimal. Both this graph and the graph in the previous exercise show that women are more likely to volunteer, but in this view, we cannot see the difference in the rate of non-participation.

Gender	Strictly voluntary	Court-ordered	Other
Men	79.16%	5.21%	15.63%
Women	85.19%	2.14%	12.67%



2.173. **(a)** At right. **(b)** $\frac{490}{800} = 61.25\%$ of male applicants are admitted, while only $\frac{400}{700} \doteq 57.14\%$ of females are admitted.

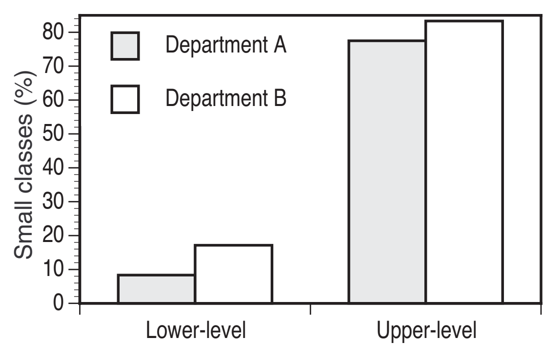
	Admit	Deny
Male	490	310
Female	400	300

(c) $\frac{400}{600} \doteq 66.67\%$ of male business school applicants are admitted; for females, this rate is the same: $\frac{200}{300} \doteq 66.67\%$. In the law school, $\frac{90}{200} = 45\%$ of males are admitted, compared to $\frac{200}{400} = 50\%$ of females. **(d)** A majority (6/7) of male applicants apply to the business school, which admits $\frac{400+200}{600+300} \doteq 66.67\%$ of all applicants. Meanwhile, a majority (3/5) of women apply to the law school, which admits only $\frac{90+200}{200+400} \doteq 48.33\%$ of its applicants.

2.174. Tables will vary, of course. The key idea is that one gender should be more likely to apply to the schools that are easier to get into. For example, if the four schools admit 50%, 60%, 70%, and 80% of applicants, and men are more likely to apply to the first two, while women apply to the latter two, women will be admitted more often.

A nice variation on this exercise is to describe two basketball teams practicing. You observe that one team makes 50% of their shots, while the other makes only 40%. Does that mean the first team is more accurate? Not necessarily; perhaps they attempted more lay-ups while the other team spent more time shooting three-pointers. (Some students will latch onto this kind of example much more quickly than discussions of male/female admission rates.)

2.175. If we ignore the "year" classification, we see that Department A teaches 32 small classes out of 52, or about 61.54%, while Department B teaches 42 small classes out of 106, or about 39.62%. (These agree with the dean's numbers.)



For the report to the dean, students may analyze the numbers in a variety of ways, some valid and some not. The key observations are: (i) When considering only first- and second-year classes, A has fewer small classes ($\frac{1}{12} \doteq 8.33\%$) than B ($\frac{12}{70} \doteq 17.14\%$). Likewise, when considering only upper-level classes, A has $\frac{31}{40} = 77.5\%$ and B has $\frac{30}{36} \doteq 83.33\%$ small classes. The graph on the right illustrates this. These numbers are given in the back of the text, so most students should include this in their analysis! (ii) $\frac{40}{52} \doteq 77.78\%$ of A's classes are upper-level courses, compared to $\frac{36}{106} \doteq 33.96\%$ of B's classes.

2.176. (a) Subtract the "agreed" counts from the sample sizes to get the "disagreed" counts. The table is in the Minitab output on the right. (The output has been slightly altered to have more descriptive row and column headings.) We find $X^2 \doteq 2.67$, df $= 1$, and $P = 0.103$, so we cannot conclude that students and non-students differ in the response to this question. **(b)** For testing H_0: $p_1 = p_2$ versus H_a: $p_1 \neq p_2$, we have $\hat{p}_1 \doteq 0.3607$, $\hat{p}_2 \doteq 0.5085$, $\hat{p} \doteq 0.4333$, $\text{SE}_{D_p} \doteq 0.09048$, and $z = -1.63$. Up to rounding, $z^2 = X^2$, and the P-values are the same. **(c)** The statistical tests in (a) and (b) assume that we have two SRSs, which we clearly do not have here. Furthermore, the two groups differed in geography (northeast/West Coast) in addition to student/non-student classification. These issues mean we should not place too much confidence in the conclusions of our significance test—or, at least, we should not generalize our conclusions too far beyond the populations "upper level northeastern college students taking a course in Internet marketing" and "West Coast residents willing to participate in commercial focus groups."

Minitab output

```
          Students   Non-st    Total
  Agr         22       30        52
           26.43    25.57

  Dis         39       29        68
           34.57    33.43

Total         61       59       120

ChiSq =  0.744 +  0.769 +
         0.569 +  0.588 = 2.669
df = 1, p = 0.103
```

2.177. (a) First we must find the counts in each cell of the two-way table. For example, there were about $(0.172)(5619) \doteq 966$ Division I athletes who admitted to wagering. These counts are shown in the Minitab output on the right, where we see that $X^2 \doteq 76.7$, df $= 2$, and $P < 0.0001$. There is very strong evidence that the percentage of athletes who admit to wagering differs by division. **(b)** Even with much smaller numbers of students (say, 1000 from each division), P is still very small. Presumably, the estimated numbers are reliable enough that we would not expect the true counts to be less than 1000, so we need not be concerned about the fact that we had to estimate the sample sizes. **(c)** If the reported proportions are wrong, then our conclusions may be suspect—especially if it is the case that athletes in some division were more likely to say they had not wagered when they had. **(d)** It is difficult to predict exactly how this might affect the results: Lack of independence could cause the estimated percents to be too large, or too small, if our sample included several athletes from teams which have (or do not have) a "gambling culture."

Minitab output

```
            Div1      Div2      Div3    Total
    1        966       621       998     2585
         1146.87    603.54    834.59

    2       4653      2336      3091    10080
         4472.13   2353.46   3254.41

Total       5619      2957      4089    12665

ChiSq = 28.525 +  0.505 + 31.996 +
         7.315 +  0.130 +  8.205 = 76.675
df = 2, p = 0.000
```

Chapter 3 Solutions

3.1. Any group of friends is unlikely to include a representative cross section of all students.

3.2. Political speeches provide a good source of examples.

3.3. A hard-core runner (and her friends) are not representative of all young people.

3.4. The performance of one car is anecdotal evidence—a "haphazardly selected individual case." People tend to remember—and re-tell—stories about extraordinary performance, while cases of average or below-average reliability are typically forgotten.

3.5. For example, who owns the Web site? Do they have data to back up this statement, and if so, what was the source of that data?

3.7. This is an experiment: Each subject is assigned to a treatment group (presumably at random, although the description does not tell us this is the case). The explanatory variable is the drug received, the response variables are adverse events, as well as some reaction. (The nature of that reaction is not specified in the exercise, but they apparently collected some information to indicate which subjects had an "effective response" to the vaccine.)

3.8. **(a)** No treatment is imposed on the subjects (children); they (or their parents) choose how much TV they watch. The explanatory variable is hours watching TV, and the response variable is "later aggressive behavior." **(b)** An adolescent who watches a lot of television probably is more likely to spend less time doing homework, playing sports, or having social interactions with peers. He or she may also have less contact with or guidance from his/her parents.

3.9. This is an experiment: Each subject is (presumably randomly) assigned to a group, each with its own treatment (computer animation or reading the textbook). The explanatory variable is the teaching method, and the response variable is the change in each student's test score.

3.10. This is an experiment, assuming the order of treatment given to each subject was randomly determined. The explanatory variable is the form of the apple (whole, or juice), and the response variable is how full the subjects felt.

Note: *This is a* matched pairs *experiment, described on page 181 of the text.*

3.11. The experimental units are food samples, the treatment is exposure to different levels of radiation, and the response variable is the amount of lipid oxidation. Note that in a study with only one factor—like this one—the treatments and factor levels are essentially the same thing: The factor is varying radiation exposure, with nine levels.

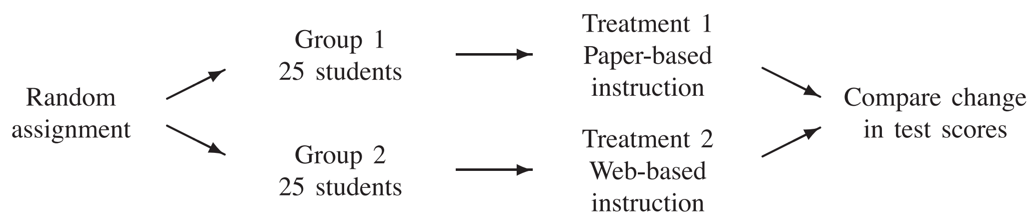
It is hard to say how much this will generalize; it seems likely that different lipids react to radiation differently.

3.12. This is an experiment because the experimental units (students) are randomly assigned to a treatment group. Note that in a study with only one factor—like this one—the treatments and factor levels are essentially the same thing: There are two treatments/levels of the factor "instruction method." The response variable is the change in score on the standardized test.

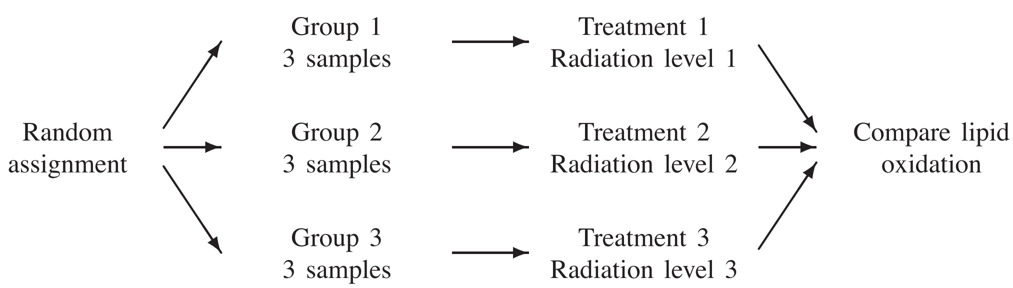
The results of this experiment should generalize to other classes (on the same topic) taught by the same instructor, but might not apply to other subject matter, or to classes taught by other instructors.

3.13. Those who volunteer to use the software may be better students (or worse). Even if we cannot decide the direction of the bias (better or worse), the lack of random allocation means that the conclusions we can draw from this study are limited at best.

3.14.



3.15. Because there are nine levels, this diagram is rather large (and repetitive), so only the top three branches are shown.



3.16. The results will depend on the software used.

3.17. **(a)** Shopping patterns may differ on Friday and Saturday, which would make it hard to determine the true effect of each promotion. (That is, the effect of the promotion would be confounded with the effect of the day.) To correct this, we could offer one promotion on a Friday, and the other on the following Friday. (Or, we could do as described in the exercise, and then on the next weekend, swap the order of the offers.) **(b)** Responses may vary in different states. To control for this, we could launch both campaigns in (separate) parts of the same state or states. **(c)** A control is needed for comparison; if we simply compare this year's yield to last year's yield, we will not know how much of the difference can be attributed to changes in the economy. We should compare the new strategy's yield with another investment portfolio using the old strategy.

Note: *For part (c), this comparison might be done without actually buying or selling anything; we could simply compute how much money would have been made if we had followed the new strategy; that is, we keep a "virtual portfolio." This assumes that our buying and selling is on a small enough scale that it does not affect market prices.*

3.18. **(a)** Assigning subjects by gender is not random. It would be better to treat gender as a blocking variable, assigning five men and five women to each treatment. **(b)** This randomization will not necessarily divide the subjects into two groups of five. (Note that it *would* be a valid randomization to use this method until one group had four subjects, and then assign any remaining subjects to the other group.) **(c)** The 10 rats in a batch might be similar to one another in some way. For example, they might be siblings, or they might have been exposed to unusual conditions during shipping. (The safest approach in this situation would be to treat each batch as a block, and randomly assign two or three rats from each batch to each treatment.)

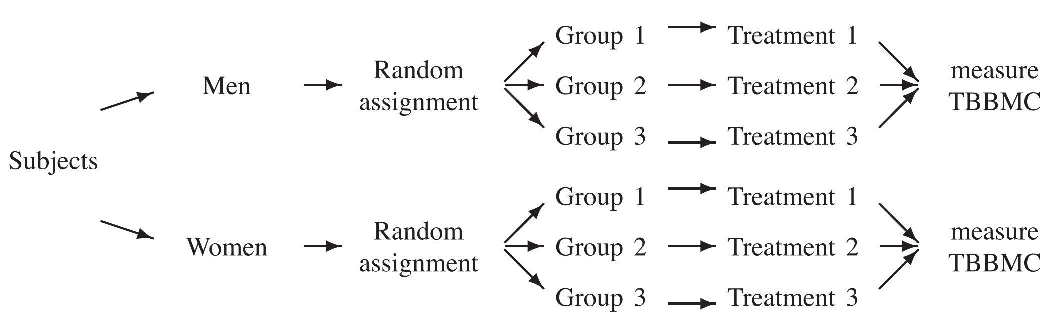
3.19. The experiment can be single-blind (those evaluating the exams should not know which teaching approach was used), but not double-blind, because the students will know which treatment (teaching method) was assigned to them.

3.20. For example, we might block by gender, by year in school, or by housing type (dorm/off-campus/Greek).

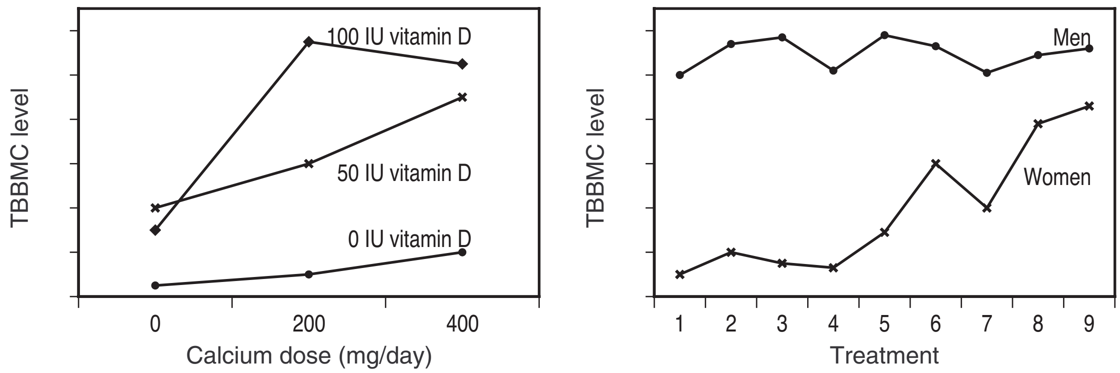
3.21. For example, new employees should be randomly assigned to either the current program or the new one. There are many possible choices for outcome variables, such as performance on a test of the information covered in the program, or a satisfaction survey or other evaluation of the program by those who went through it.

3.22. Subjects—perhaps recruited from people suffering from chronic pain, or those recovering from surgery or an injury—should be randomly assigned to be treated with magnets, or a placebo (an object similar to the magnets, except that it is not magnetic. Students should address some of the practical difficulties of such an experiment, such as: How does one measure pain relief? How can we prevent subjects from determining whether they are being treated with a magnet? (For the latter question, we might apply the treatments in a controlled setting, making sure that there is nothing metal with which the subjects could test their treatment object.)

3.23. **(a)** The factors are calcium dose, and vitamin D dose. There are 9 treatments (each calcium/vitamin D combination). **(b)** Assign 20 students to each group, with 10 of each gender. The complete diagram (including the blocking step) would have a total of 18 branches, below is a portion of that diagram, showing only three of the nine branches for each gender. **(c)** Randomization results will vary.



3.24. Students may need guidance as to how to illustrate these interactions. Figure 3.7 shows one such illustration (as part of Exercise 3.40). Shown below are two possible illustrations, based on made-up data (note there is no scale on the vertical axis). In the first, we see that the effect of vitamin D on TBBMC depends on the calcium dose. In the second, we see little variation in men's TBBMC across all nine treatments, while women's TBBMC appears to depend on the treatment group. (In particular, women's TBBMC is greatest for treatment 9, with the highest doses of both calcium and vitamin D.)



3.25. **(a)** For example, flip a coin for each customer to choose which variety (s)he will taste. To evaluate preferences, we would need to design some scale for customers to rate the coffee they tasted, and then compare the ratings. **(b)** For example, flip a coin for each customer to choose which variety (s)he will taste *first*. Ask which of the two coffees (s)he preferred. **(c)** The matched-pairs version described in part (b) is the stronger design; if each customer tastes both varieties, we only need to ask which was preferred. In part (a), we might ask customers to rate the coffee they tasted on a scale of (say) 1 to 10, but such ratings can be wildly variable (one person's "5" might be another person's "8"), which makes comparison of the two varieties more difficult.

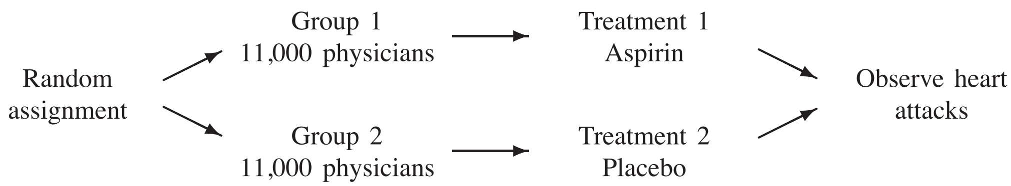
3.26. An experiment would be nearly impossible in this case, unless the publishers of *Sports Illustrated* would allow you to choose the cover photo for several issues. (The ideal design would involve taking a collection of currently-successful teams or individuals, and randomly assigning half to be on the cover, then observing their fortunes over the next few months.)

3.27. Experimental units: pine tree seedlings. Factor: amount of light. Treatments (two): full light, or shaded to 5% of normal. Response variable: dry weight at end of study.

3.28. Experimental units: Middle schools. Factors: Physical activity program, and nutrition program. Treatments (four): Activity intervention, nutrition intervention, both interventions, and neither intervention. Response variables: Physical activity and lunchtime consumption of fat.

3.29. Subjects: adults (or registered voters) from selected households. Factors: level of identification, and offer of survey results. Treatments (six): interviewer's name with results, interviewer's name without results, university name with results, university name without results, both names with results, both names without results. Response variable: whether or not the interview is completed.

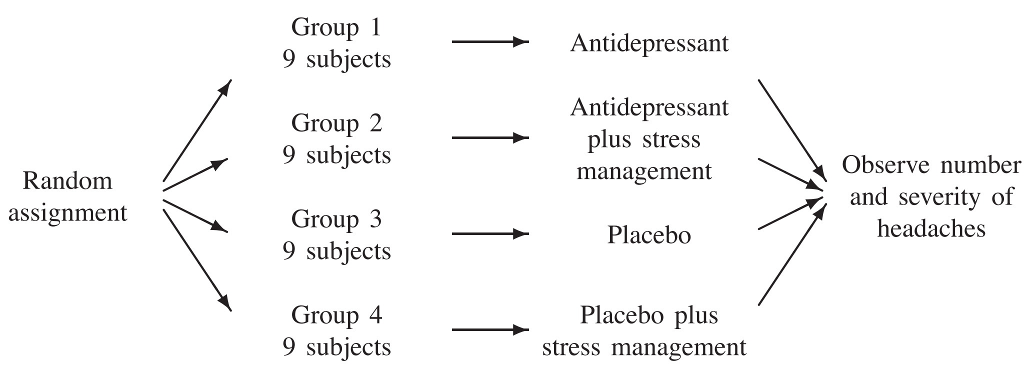
3.30. **(a)** The subjects are the physicians, the factor is medication (aspirin or placebo), and the response variable is health, specifically whether the subjects have heart attacks. **(b)** Below. **(c)** The difference in the number of heart attacks between the two groups was so great that it would rarely occur by chance if aspirin had no effect.



3.31. Assign nine subjects to treatment 1, then nine more to treatment 2, etc. A diagram is on the next page; if we assign labels 01 through 36, then line 151 gives:

Group 1		Group 2		Group 3	
03 Bezawada	12 Hatfield	32 Tyner	27 Rau	05 Cheng	13 Hua
22 Mi	11 Guha	30 Tang	20 Martin	16 Leaf	25 Park
29 Shu	31 Towers	09 Daye	06 Chronopoulou	28 Saygin	19 Lu
26 Paul	21 Mehta	23 Nolan	33 Vassilev	10 Engelbrecht	04 Cetin
01 Anderson		07 Codrington		18 Lipka	

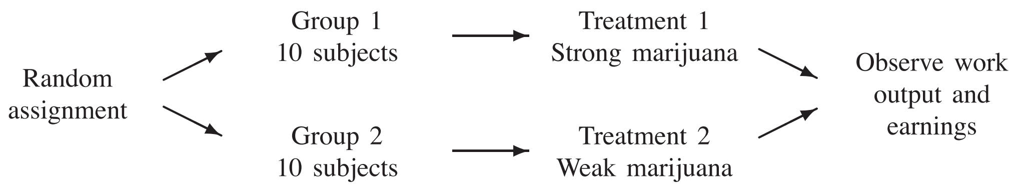
The other nine subjects are in Group 4. The names listed here assume that labels are assigned alphabetically (across the rows). See note on page 50 about using Table B.



3.32. (a) A diagram is shown below. **(b)** Label the subjects from 01 through 20. From line 101, we choose:

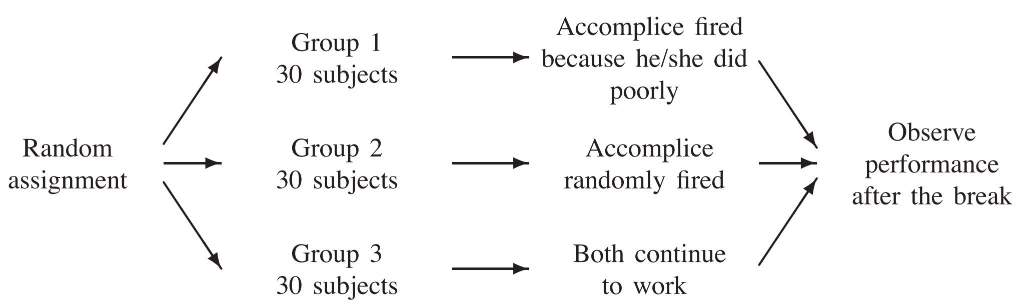
19, 05, 13, 17, 09, 07, 02, 01, 18, and 14

Assuming that labels are assigned alphabetically, that is Wayman, Cunningham, Mitchell, Seele, Knapp, Fein, Brifcani, Becker, Truong, and Ponder for one group, and the rest for the other. See note on page 50 about using Table B.

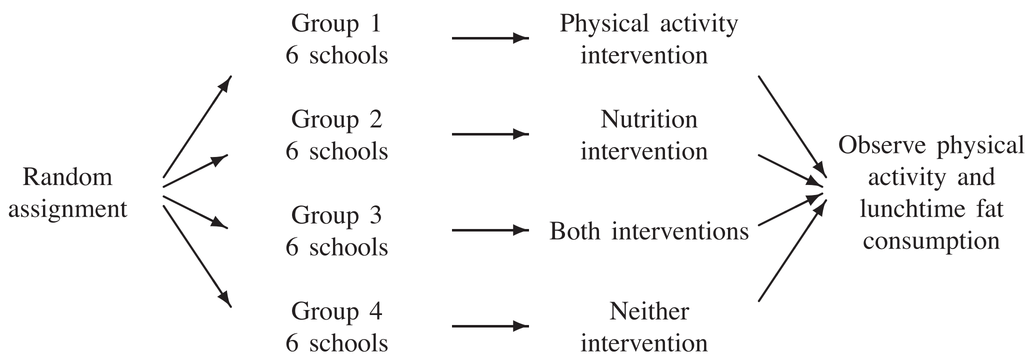


3.33. Students might envision different treatments; one possibility is to have some volunteers go through a training session, while others are given a written set of instructions, or watch a video. For the response variable(s), we need some measure of training effectiveness; perhaps we could have the volunteers analyze a sample of lake water and compare their results to some standard.

3.34. (a) Diagram below. **(b)** Using line 153 from Table B, the first four subjects are 07, 88, 65, and 68. See note on page 50 about using Table B.



3.35. Diagram below. Starting at line 160, we choose:
16, 21, 06, 12, 02, 04 for Group 1
14, 15, 23, 11, 09, 03 for Group 2
07, 24, 17, 22, 01, 13 for Group 3
The rest are assigned to Group 4. See note on page 50 about using Table B.



3.36. **(a)** The table below shows the 16 treatments—four levels for each of the two factors. **(b)** A diagram is not shown here (with 16 treatments, it is quite large). Six subjects are randomly assigned to each treatment; they read the ad for that treatment, and we record their attractiveness ratings for the ad. Using line 111, the first six subjects are 81, 48, 66, 94, 87, and 60.

		Factor B Fraction of shoes on sale			
		25%	50%	75%	100%
Factor A Discount level	20%	1	2	3	4
	40%	5	6	7	8
	60%	9	10	11	12
	80%	13	14	15	16

3.37. **(a)** Population = 1 to **150**, Select a sample of size **25**, click Reset and Sample. **(b)** Without resetting, click Sample again. **(c)** Click Sample three more times.

3.38. Population = 1 to **40**, Select a sample of size **20**, then click Reset and Sample.

3.39. Design (a) is an experiment. Because the treatment is randomly assigned, the effect of other habits would be "diluted" because they would be more-or-less equally split between the two groups. Therefore, any difference in colon health between the two groups could be attributed to the treatment (bee pollen or not).

Design (b) is an observational study. It is flawed because the women observed chose whether or not to take bee pollen; one might reasonably expect that people who choose to take bee pollen have other dietary or health habits that would differ from those who do not.

3.40. For a range of discounts, the attractiveness of the sale decreases slightly as the percentage of goods on sale increases. (The decrease is so small that it might not be significant.) With precise discounts, on the other hand, mean attractiveness increases with the percentage on sale. Range discounts are more attractive when only 25% of goods are marked down, while the precise discount is more attractive if 75% or 100% of goods are discounted.

3.41. As described, there are two factors: ZIP code (three levels: none, five-digit, nine-digit) and the day on which the letter is mailed (three levels: Monday, Thursday, or Saturday) for a total of nine treatments. To control lurking variables, aside from mailing all letters to the same address, all letters should be the same size, and either printed in the same handwriting or typed. The design should also specify how many letters will be in each treatment group. Also, the letters should be sent randomly over many weeks.

3.42. Results will vary, but probability computations reveal that more than 97% of samples will have 7 to 13 fast-reacting subjects (and 99.6% of samples have 8 to 14 fast-reacting subjects). Additionally, if students average their 10 samples, nearly all students (more than 99%) should find that the average number of fast-reacting subjects is between 8.5 and 11.5.

Note: *X, the number of fast-reacting subjects in the sample, has a hypergeometric distribution with parameters* $N = 40$, $r = 20$, $n = 20$, *so that* $P(7 \leq X \leq 13) \doteq 0.974$. *The theoretical average number of fast-reacting subjects is 10.*

3.43. Each player will be put through the sequence (100 yards, four times) twice—once with oxygen and once without, and we will observe the difference in their times on the final run. (If oxygen speeds recovery, we would expect that the oxygen-boosted time will be lower.) Randomly assign half of the players to use oxygen on the first trial, while the rest use it on the second trial. Trials should be on different days to allow ample time for full recovery.

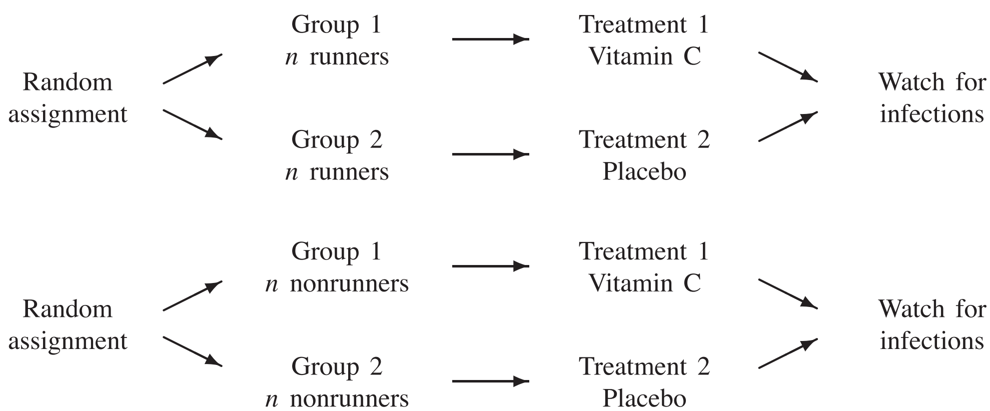
If we label the players 01 through 20 and begin on line 140, we choose 12, 13, 04, 18, 19, 16, 02, 08, 17, 10 to be in the oxygen-first group. See note on page 50 about using Table B.

3.44. The sketches requested in the problem are not shown here; random assignments will vary among students. **(a)** Label the circles 1 to 6, then randomly select three (using Table B, or simply by rolling a die) to receive the extra CO_2. Observe the growth in all six regions, and compare the mean growth within the three treated circles with the mean growth in the other three (control) circles. **(b)** Select pairs of circles in each of three different areas of the forest. For each pair, randomly select one circle to receive the extra CO_2 (using Table B or by flipping a coin). For each pair, compute the difference in growth (treated minus control).

3.45. **(a)** Randomly assign half the girls to get high-calcium punch; the other half will get low-calcium punch. The response variable is not clearly described in this exercise; the best we can say is "observe how the calcium is processed." **(b)** Randomly select half of the girls to receive high-calcium punch first (and low-calcium punch later), while the other half gets low-calcium punch first (followed by high-calcium punch). For each subject, compute the difference in the response variable for each level. This is a better design because it deals with person-to-person variation; the differences in responses for 60 individuals gives more precise results than the difference in the average responses for two groups of 30 subjects. **(c)** The first five subjects are 38, 44, 18, 33, and 46. In the CR design, the first group receives high-calcium punch all summer; in the matched pairs design, they receive high-calcium punch for the first part of the summer, and then low-calcium punch in the second half.

3.46. **(a)** False. Such regularity holds only in the long run. If it were true, you could look at the first 39 digits and know whether or not the 40th was a 0. **(b)** True. All pairs of digits (there are 100, from 00 to 99) are equally likely. **(c)** False. Four random digits have chance 1/10000 to be 0000, so this sequence will occasionally occur. 0000 is no more or less random than 1234 or 2718 or any other four-digit sequence.

3.47. **(a)** This is a block design. **(b)** The diagram might be similar to the one below (which assumes equal numbers of subjects in each group). **(c)** The results observed in this study would rarely have occurred by chance if vitamin C were ineffective.



3.48. The population is faculty members at Mongolian public universities, and the sample is the 300 faculty members to whom the survey was sent. Because we do not know how many responses were received, we cannot determine the response rate.

Note: *We might consider the population to be either the 2500 faculty members on the list, or the larger group of "all current and future faculty members." In the latter case, those on the list constitute the* sampling frame—*the subset of the population from which our sample will be selected.*

The sample might be considered to be only those faculty who actually responded to the survey (rather than the 300 selected), because that is the actual group from which we "draw conclusions about the whole."

3.49. The population is all forest owners in the region. The sample is the 772 forest owners contacted. The response rate is $\frac{348}{772} \doteq 45\%$. Aside from the given information, we would like to know the sample design (and perhaps some other things).

Note: *It would also be reasonable to consider the sample to be the 348 forest owners who returned the survey, because that is the actual group from which we "draw conclusions about the whole."*

3.50. To use Table B, number the list from 0 to 9 and choose two single digits. (One can also assign labels 01–10, but that would require two-digit numbers, and we would almost certainly end up skipping over many pairs of digits before we found two in the desired range.)

It is worth noting that choosing an SRS is often described as "pulling names out of a hat." For long lists, it is often impractical to do this literally, but with such a small list, one really could write each ringtone on a slip of paper and choose two slips at random.

3.51. See the solution to the previous exercise; for this problem, we need to choose three items instead of two, but it is otherwise the same.

3.52. **(a)** This is a multistage sample: We first sample three of the seven course sections, then eight from each chosen section. **(b)** This is an SRS: Each student has the same chance (5/55) of being selected. **(c)** This is a voluntary response sample: Only those who visit the site can participate (if they choose to). **(d)** This is a stratified random sample: Males and females (the strata) are sampled separately.

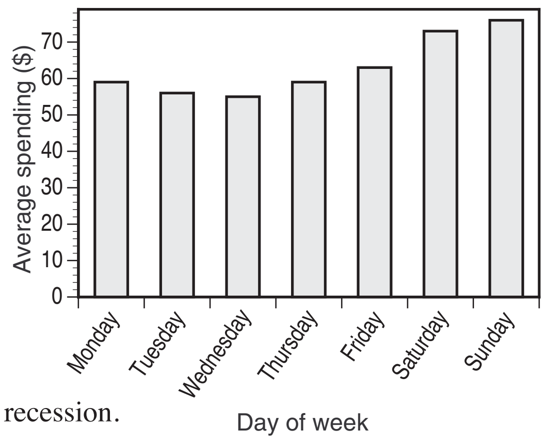
3.53. **(a)** This statement confuses the ideas of population and sample. (If the entire population is found in our sample, we have a *census* rather than a sample.) **(b)** "Dihydrogen monoxide" is H_2O. Any concern about the dangers posed by water most likely means that the respondent did not know what dihydrogen monoxide was, and was too embarrassed to admit it. (Conceivably, the respondent knew the question was about water and had concerns arising from a bad experience of flood damage or near-drowning. But misunderstanding seems to be more likely.) **(c)** Honest answers to such questions are difficult to obtain even in an anonymous survey; in a public setting like this, it would be surprising if there were any raised hands (even though there are likely to be at least a few cheaters in the room).

3.54. **(a)** The content of a single chapter is not random; choose random words from random pages. **(b)** Students who are registered for an early-morning class might have different characteristics from those who avoid such classes. **(c)** Alphabetic order is not random. One problem is that the sample might include people with the same last name (siblings, spouses, etc.). Additionally, some last names tend to be more common in some ethnic groups.

3.55. The population is (all) local businesses. The sample is the 73 businesses that return the questionnaire, *or* the 150 businesses selected. The nonresponse rate is $51.3\% = \frac{77}{150}$.

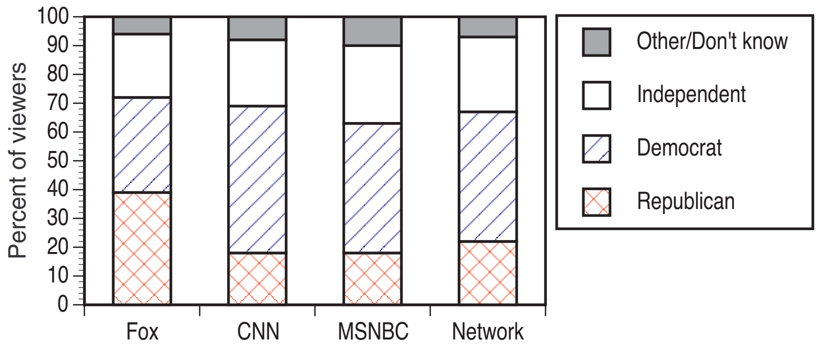
Note: *The definition of "sample" makes it somewhat unclear whether the sample includes all the businesses selected or only those that responded. My inclination is toward the latter (the smaller group), which is consistent with the idea that the sample is "a part of the population that we actually examine."*

3.56. **(a)** Shown is a possible display (as a bar graph). A plot similar to Figure 3.7 would tell the same story—namely, that spending is highest on the weekends, and drops during the middle of the week. Ideally, bars should be in chronological order, although students might choose to start with Sunday rather than Monday. Some students might arrange bars in increasing or decreasing order of height. **(b)** The exclusion of the Christmas shopping season might impact the numbers. In addition, much of this period fell during a economic recession.



3.57. Note that the numbers add to 100% down the columns; that is, 39% is the percent of Fox viewers who are Republicans, *not* the percent of Republicans who watch Fox.

Students might display the data using a stacked bar graph like the one below, or side-by-side bars. (They could also make four pie charts, but comparing slices across pies is difficult.) The most obvious observation is that the party identification of Fox's audience is noticeably different from the other three sources.



3.58. Exact descriptions of the populations may vary. **(a)** All current students—or perhaps all current students who were enrolled during the year prior to the change. (The latter would be appropriate if we want opinions based on a comparison between the new and old curricula.) **(b)** All U.S. households. **(c)** Adult residents of the United States.

3.59. Numbering from 01 to 33 alphabetically (down the columns), we enter Table B at line 137 and choose:

12 = Country View, 14 = Crestview, 11 = Country Squire, 16 = Fairington, 08 = Burberry

See note on page 50 about using Table B.

3.60. Assign labels 001 to 200. To use Table B, take three digits at a time beginning on line 112; the first five pixels are 089, 064, 032, 117, and 003.

3.61. Population = 1 to **200** , Select a sample of size **25** , then click Reset and Sample .

3.62. With the applet: Population = 1 to **371**, Select a sample of size **25**, then click Reset and Sample. With Table B, line 120 gives the codes labeled 354, 239, 193, 099, and 262.

3.63. One could use the labels already assigned to the blocks, but that would mean skipping a lot of four-digit combinations that do not correspond to any block. An alternative would be to drop the second digit and use labels 100–105, 200–211, and 300–325. But by far the simplest approach is to assign labels 01–44 (in numerical order by the four-digit numbers already assigned), enter the table at line 135, and select:

39 (block 3020), 10 (2003), 07 (2000), 11 (2004), and 20 (3001)

See note on page 50 about using Table B.

3.64. If one always begins at the same place, then the results could not really be called random.

3.65. The sample will vary with the starting line in Table B. The simplest method is to use the last digit of the numbers assigned to the blocks in Group 1 (that is, assign the labels 0–5), then choose one of those blocks; use the last two digits of the blocks in Group 2 (00–11) and choose two of those, and finally use the last two digits of the blocks in Group 3 (00–25) and choose three of them.

3.66. **(a)** If we choose one of the first 45 students and then every 45th name after that, we will have a total of $\frac{9000}{45} = 200$ names. **(b)** Label the first 45 names 01–45. Beginning at line 125, the first number we find is 21, so we choose names 21, 66, 111, ...

3.67. Considering the 9000 students of Exercise 3.66, each student is equally likely; specifically, each name has chance 1/45 of being selected. To see this, note that each of the first 45 has chance 1/45 because one is chosen at random. But each student in the second 45 is chosen exactly when the corresponding student in the first 45 is, so each of the second 45 also has chance 1/45. And so on.

This is not an SRS because the only possible samples have exactly one name from the first 45, one name from the second 45, and so on; that is, there are only 45 possible samples. An SRS could contain *any* 200 of the 9000 students in the population.

3.68. **(a)** This is a stratified random sample. **(b)** Label from 01 through 27; beginning at line 122, we choose:

13 (805), 15 (760), 05 (916), 09 (510), 08 (925),
27 (619), 07 (415), 10 (650), 25 (909), and 23 (310)

Note: *The area codes are in north-south order if we read across the rows; that is how they were labeled for this solution. Students might label down rather than across; the sample should include the same set of labels but a different list of area codes.*

3.69. Assign labels 01–36 for the Climax 1 group, 01–72 for the Climax 2 group, and so on. Then beginning at line 140, choose:

12, 32, 13, 04 from the Climax 1 group and (continuing on in Table B)
51, 44, 72, 32, 18, 19, 40 from the Climax 2 group
24, 28, 23 from the Climax 3 group and
29, 12, 16, 25 from the mature secondary group

See note on page 50 about using Table B.

3.70. Label the students 01, . . . , 30 and use Table B. Then label the faculty 0, . . . , 9 and use the table again. (You could also label the faculty from 01 to 10, but that would needlessly require two-digit labels.)

Note: *Students often try some fallacious method of choosing both samples simultaneously. We simply want to choose two separate SRSs: one from the students and one from the faculty. See note on page 50 about using Table B.*

3.71. Each student has a 10% chance: 3 out of 30 over-21 students, and 2 of 20 under-21 students. This is not an SRS because not every group of 5 students can be chosen; the only possible samples are those with 3 older and 2 younger students.

3.72. Label the 500 midsize accounts from 001 to 500, and the 4400 small accounts from 0001 to 4400. On line 115, we first encounter numbers 417, 494, 322, 247, and 097 for the midsize group, then 3698, 1452, 2605, 2480, and 3716 for the small group. See note on page 50 about using Table B.

3.73. **(a)** This design would omit households without telephones or with unlisted numbers. Such households would likely be made up of poor individuals (who cannot afford a phone), those who choose not to have phones, and those who do not wish to have their phone numbers published. **(b)** Those with unlisted numbers would be included in the sampling frame when a random-digit dialer is used.

3.74. **(a)** This will almost certainly produce a positive response because it draws the dubious conclusion that cell phones *cause* brain cancer. Some people who drive cars, or eat carrots, or vote Republican develop brain cancer, too. Do we conclude that these activities should come with warning labels, also? **(b)** The phrasing of this question will tend to make people respond in favor of national health insurance: It lists two benefits of such a system, and no arguments from the other side of the issue. **(c)** This sentence is so convoluted and complicated that it is almost unreadable; it is also vague (what sort of "economic incentives"? How much would this cost?). A better phrasing might be, "Would you be willing to pay more for the products you buy if the extra cost were used to conserve resources by encouraging recycling?" That is still vague, but less so, and is written in plain English.

3.75. The first wording brought the higher numbers in favor of a tax cut; "new government programs" has considerably less appeal than the list of specific programs given in the second wording.

3.76. Children from larger families will be overrepresented in such a sample. Student explanations of why will vary; a simple illustration can aid in understanding this effect. Suppose that there are 100 families with children; 60 families have one child and the other 40 have three. Then there are a total of 180 children (an average of 1.8 children per family), and *two-thirds* (120) of those children come from families with three children. Therefore, if we had a class (a sample) chosen from these 180 children, only one-third of the class would answer "one" to the teacher's question, and the rest would say "three." This would give an average of about 2.3 children per family.

3.78. Responses to public opinion polls can be affected by things like the wording of the question, as was the case here: Both statements address the question of how to distribute wealth in a society, but subtle (and not-so-subtle) slants in the wording suggest that the public holds conflicting opinions on the subjects.

3.79. The population is undergraduate college students. The sample is the 2036 students. (We assume they were randomly selected.)

3.80. No; this is a voluntary response sample. The procedures described in the text apply to data gathered from an SRS.

3.81. The larger sample would have less sampling variability. (That is, the results would have a higher probability of being closer to the "truth.")

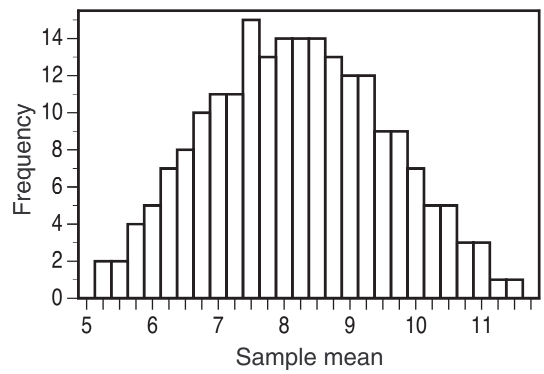
3.82. **(a)** Parameters are associated with the population; statistics describe samples. **(b)** Bias means that the center of sampling distribution is not equal to the true value of the parameter; that is, bias is systematic under- or over-estimation. Variability refers to the spread (not the center) of the sampling distribution. **(c)** Large samples generally have lower variability, but if the samples are biased, that lower variability is of little use. (In addition, larger samples generally come at a cost; the added cost might not justify the decrease in variability.) **(d)** A sampling distribution might be visualized (or even simulated) with a computer, but it arises from the process of sampling, not from computation.

3.83. **(a)** Population: college students. Sample: 17,096 students. **(b)** Population: restaurant workers. Sample: 100 workers. **(c)** Population: longleaf pine trees. Sample: 40 trees.

3.84. **(a)** High bias, high variability (many are low, wide scatter). **(b)** Low bias, low variability (close to parameter, little scatter). **(c)** Low bias, high variability (neither too low nor too high, wide scatter). **(d)** High bias, low variability (too high, little scatter).

Note: *Make sure that students understand that "high bias" means that the values are far from the parameter,* not *that they are too high.*

3.85. **(a)** The scores will vary depending on the starting row. Note that the smallest possible mean is 5.25 (from the sample 3, 5, 6, 7) and the largest is 11.5 (from 9, 10, 12, 15). **(b)** Answers will vary. On the right is the (exact) sampling distribution, showing all possible values of the experiment (so the first rectangle is for 5.25, the next is for 5.5, etc.). Note that it looks roughly Normal; if we had taken a larger sample from a larger population, it would appear even more Normal.



Note: *This histogram was found by considering all* $\binom{10}{4} = 210$ *of the possible samples. A collection of only 10 random samples will, of course, be considerably less detailed.*

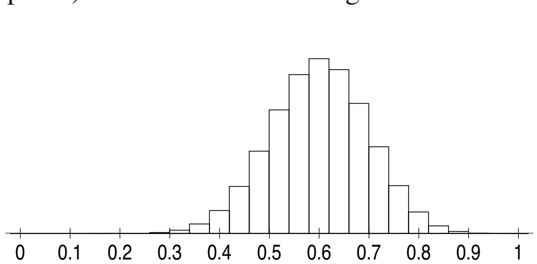
3.86. No: With sufficiently large populations ("at least 100 times larger than the sample"), the variability (and margin of error) depends on the sample size.

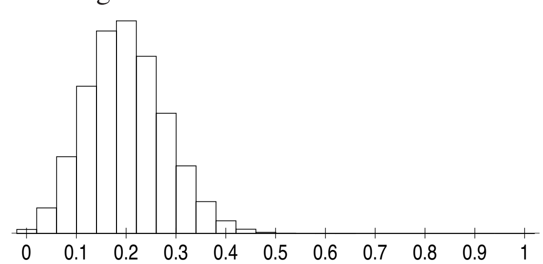
3.87. **(a)** This is a multistage sample. **(b)** Attitudes in smaller countries (many of which were not surveyed) might be different. **(c)** An individual country's reported percent will typically differ from its true percent by no more than the stated margin of error. (The margins of error differ among the countries because the sample sizes were not all the same.)

Note: *The number of countries in the world is about 195 (the exact number depends on the criteria of what constitutes a separate country). That means that about 60 countries are not represented in this survey.*

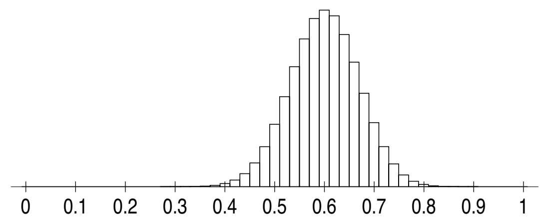
3.88. **(a)** The population is Ontario residents; the sample is the 61,239 people interviewed. **(b)** The sample size is very large, so if there were large numbers of both sexes in the sample—this is a safe assumption because we are told this is a "random sample"—these two numbers should be fairly accurate reflections of the values for the whole population.

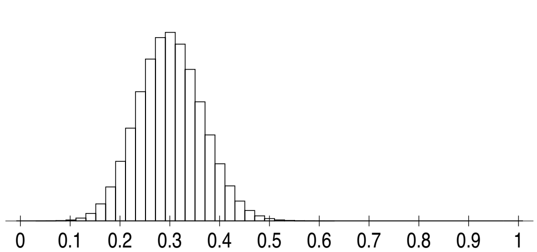
3.89. **(a)** The histogram should be centered at about 0.6 (with quite a bit of spread). For reference, the theoretical histogram is shown below on the left; student results should have a similar appearance. **(b)** The histogram should be centered at about 0.2 (with quite a bit of spread). The theoretical histogram is shown below on the right.

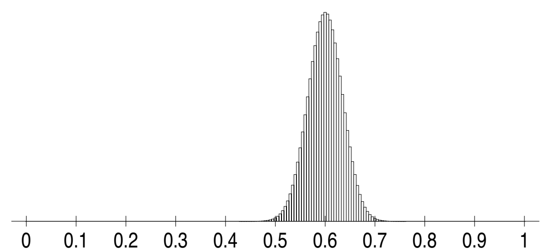




3.90. **(a)** The histogram of this theoretical sampling distribution is shown (on the right) for reference. **(b)** This theoretical sampling distribution is shown below on the left. Students should observe that their two stemplots have clearly different centers (near 0.6 and 0.3, respectively) but similar spreads. **(c)** The theoretical sampling distribution is below on the right. Compared to the distribution of part (a), this has the same center but is about half as wide; that is, the spread is about half as much when the sample size is multiplied by 4. (The vertical scale of this graph is not the same as the other two; it should be about twice as tall as it is since it is only about half as wide.)

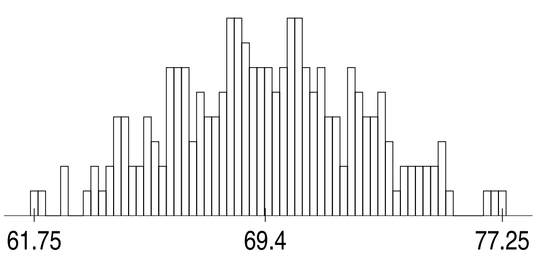


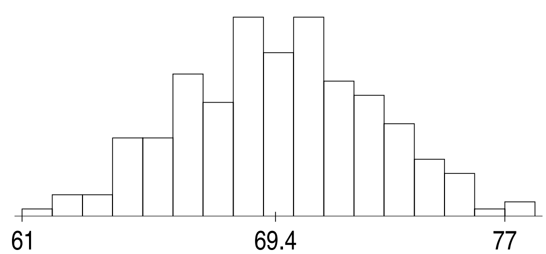




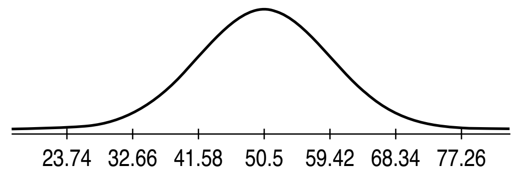
3.91. **(a)** The scores will vary depending on the starting row. Note that the smallest possible mean is 61.75 (from the sample 58, 62, 62, 65) and the largest is 77.25 (from 73, 74, 80, 82). **(b)** Answers will vary; shown below are two views of the (exact) sampling distribution. The first shows all possible values of the experiment (so the first rectangle is for 61.75, the next is for 62.00, etc.); the other shows values grouped from 61 to 61.75, 62 to 62.75, etc. (which makes the histogram less bumpy). The tallest rectangle in the first picture is 8 units; in the second, the tallest is 28 units.

Note: *These histograms were found by considering all* $\binom{10}{4} = 210$ *of the possible samples. It happens that half (105) of those samples yield a mean smaller than 69.4 and half yield a greater mean.*

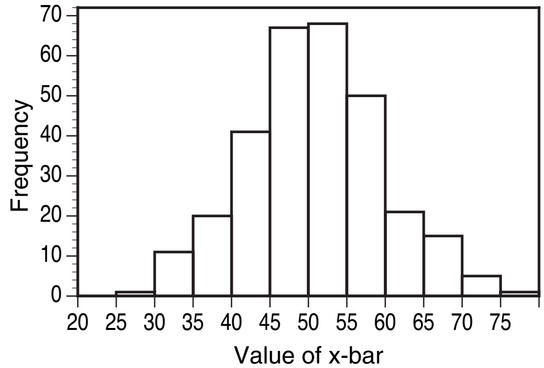




3.92. Student results will vary greatly, and ten values of $\bar{x}$ will give little indication of the appearance of the sampling distribution. In fact, the sampling distribution of $\bar{x}$ is approximately Normal with a mean of 50.5 and a standard deviation of about 8.92; this approximating Normal distribution is shown on the right (above). Therefore, nearly every sample of size 10 would yield a mean between 23 and 78.



The shape of the sampling distribution becomes more apparent if the results of many students are pooled. Below on the right is an example based on 300 sample means, which might arise from pooling all the results in a class of 30.

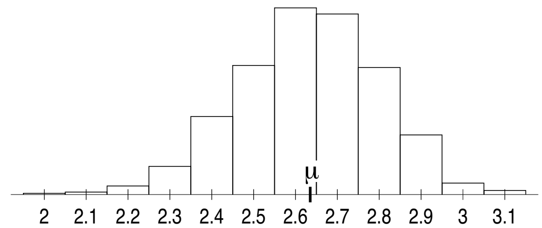


Note: *Because the values in these samples are not independent (there can be no repeats), a stronger version of the central limit theorem is needed to determine that the sampling distribution is approximately Normal. Confirming the standard deviation given above is a reasonably difficult exercise even for a mathematics major.*

3.93. (a) Below is the population stemplot (which gives the same information as a histogram). The (population) mean GPA is $\mu \doteq 2.6352$, and the standard deviation is $\sigma \doteq 0.7794$. [Technically, we should take $\sigma \doteq 0.7777$, which comes from dividing by n rather than $n - 1$, but few (if any) students would know this, and it has little effect on the results.]
(b) & **(c)** Results will vary; these histograms are not shown. Not every sample of size 20 could be viewed as "generally representative of the population," but most should bear at least some resemblance to the population distribution.

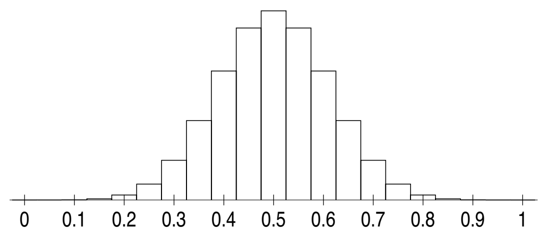
0	134
0	567889
1	0011233444
1	5566667888888888999999
2	000000000111111111222222222233333333444444444
2	55555555555556666666667777777777777788888888888888999999
3	000000000000011111111112222222222333333333333333444444444
3	5566666666677777788889
4	0000

3.94. **(a)** Shown for reference is a histogram of the approximate sampling distribution of $\bar{x}$. This distribution is difficult to find exactly, but based on 1000 simulated samples, it is approximately Normal with mean 2.6352 (the same as μ) and standard deviation $s_{\bar{x}} \doteq 0.167$. (Therefore, $\bar{x}$ will almost always be between 2.13 and 3.14.) **(b)** Results may vary, but most students should see no strong suggestion of bias. **(c)** Student means and standard deviations will vary, but for most (if not all) students, their values should meet the expectations (close to $\mu \doteq 2.6352$ and less than $\sigma \doteq 0.78$).



Note: *Observe that the distribution of $\bar{x}$ is slightly left-skewed, but less skewed than the population distribution. Also note that $s_{\bar{x}}$, the standard deviation of the sampling distribution, is smaller than $\sigma/\sqrt{20} \doteq 0.174$, since we are sampling without replacement.*

3.95. **(a)** Answers will vary. If, for example, eight heads are observed, then $\hat{p} = \frac{8}{20} = 0.4 = 40\%$. **(b)** Note that all the leaves in the stemplot should be either 0 or 5 since all possible $\hat{p}$-values end in 0 or 5. For comparison, here is a histogram of the sampling distribution (assuming p really is 0.5). An individual student's stemplot will probably only roughly approximate this distribution, but pooled efforts should be fairly close.



Many of the questions in Section 3.4 (Ethics), Exercises 3.96–3.117, are matters of opinion and may be better used for class discussion rather than as assigned homework. A few comments are included here.

3.96. These three proposals are clearly in increasing order of risk. Most students will likely consider that (a) qualifies as minimal risk, and most will agree that (c) goes beyond minimal risk.

3.97. **(a)** A nonscientist might raise different viewpoints and concerns from those considered by scientists. **(b)** Answers will vary.

3.98. It is good to plainly state the purpose of the research ("To study how people's religious beliefs and their feelings about authority are related"). Stating the research *thesis* (that orthodox religious belief are associated with authoritarian personalities) would cause bias.

3.102. **(a)** Ethical issues include informed consent and confidentiality; random assignment generally is not an ethical consideration. **(b)** "Once research begins, the board monitors its progress at least once a year." **(c)** Harm need not be physical; psychological harm also needs to be considered.

3.103. To control for changes in the mass spectrometer over time, we should alternate between control and cancer samples.

3.105. The articles are "Facebook and academic performance: Reconciling a media sensation with data" (Josh Pasek, eian more, Eszter Hargittai), a critique of the first article called "A response to reconciling a media sensation with data" (Aryn C. Karpinski), and a response to the critique ("Some clarifications on the Facebook-GPA study and Karpinski's response") by the original authors. In case these articles are not available at the address given in the text, they might be found elsewhere with a Web search.

3.106. **(a)** The simplest approach is to label from 00001 through 14959 and then take five digits at a time from the table. A few clever students might think of some ways to make this process more efficient, such as taking the first random digit chosen as "0" if it is even and "1" if odd. (This way, fewer numbers need to be ignored.) **(b)** Using labels 00001–14959, we choose 05995, 06788, and 14293. Students who try an alternate approach may have a different sample.

3.108. **(a)** A sample survey: We want to gather information about a population (U.S. residents) based on a sample. **(b)** An experiment: We want to establish a cause-and-effect relationship between teaching method and amount learned. **(c)** An observational study: There is no particular population from which we will sample; we simply observe "your teachers," much like an animal behavioral specialist might study animals in the wild.

3.111. They cannot be anonymous because the interviews are conducted in person in the subject's home. They are certainly kept confidential.

Note: *For more information about this survey, see the GSS Web site:*
`www.norc.org/GSS+Website`

3.112. This offers anonymity, since names are never revealed. (However, faces are seen, so there may be some chance of someone's identity becoming known.)

3.116. **(a)** Those being surveyed should be told the kind of questions they will be asked and the approximate amount of time required. **(b)** Giving the name and address of the organization may give the respondents a sense that they have an avenue to complain should they feel offended or mistreated by the pollster. **(c)** At the time that the questions are being asked, knowing who is paying for a poll may introduce bias, perhaps due to nonresponse (not wanting to give what might be considered a "wrong" answer). When information about a poll is made public, though, the poll's sponsor should be announced.

3.120. At `norc.org`, search for "Consumer Finances" or "SCF," and from the SCF page, click on the link to "website for SCF respondents." At the time this manual was written, the pledge was found at `www.norc.org/scf2010/Confidentiality.html`.

3.121. **(a)** You need information about a random selection of his games, not just the ones he chooses to talk about. **(b)** These students may have chosen to sit in the front; all students should be randomly assigned to their seats.

3.122. **(a)** A matched pairs design (two halves of the same board would have similar properties). **(b)** A sample survey (with a stratified sample: smokers and nonsmokers). **(c)** A block design (blocked by gender).

3.123. This is an experiment because each subject is (randomly, we assume) assigned to a treatment. The explanatory variable is the price history seen by the subject (steady prices or fluctuating prices), and the response variable is the price the subject expects to pay.

3.124. **(a)** For example, one could select (or recruit) a sample and assess each person's calcium intake (perhaps by having them record what they eat for a week), and measure his/her TBBMD. **(b)** For example, measure each subject's TBBMD, then randomly assign half the subjects to take a calcium supplement, and the other half to take a placebo. After a suitable period, measure TBBMD again. **(c)** The experiment, while more complicated, gives better information about the relationship between these variables, because it controls for other factors that may affect bone health.

3.126. Each subject should taste both kinds of fries in a randomly selected order and then be asked about preference. One question to consider is whether they should have ketchup available; many people typically eat fries with ketchup, and its presence or absence might affect their preferences. If ketchup is used, should one use the same ketchup for both, or a sample of the ketchup from each restaurant?

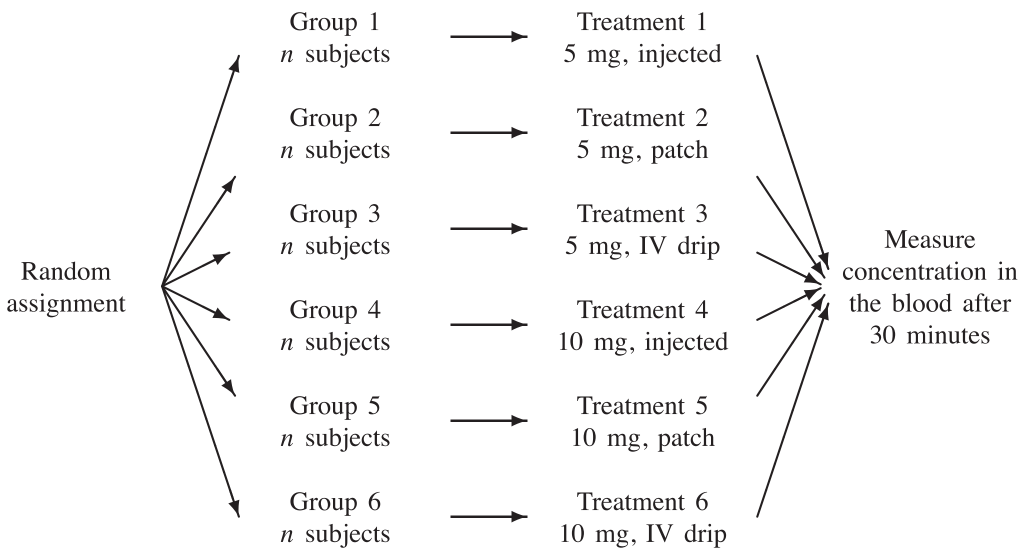
3.127. The two factors are gear (three levels) and steepness of the course (number of levels not specified). Assuming there are at least three steepness levels—which seems like the smallest reasonable choice—that means at least nine treatments. Randomization should be used to determine the order in which the treatments are applied. Note that we must allow ample recovery time between trials, and it would be best to have the rider try each treatment several times.

3.129. **(a)** One possible population: all full-time undergraduate students in the fall term on a list provided by the registrar. **(b)** A stratified sample with 125 students from each year is one possibility. **(c)** Mailed (or emailed) questionnaires might have high nonresponse rates. Telephone interviews exclude those without phones and may mean repeated calling for those who are not home. Face-to-face interviews might be more costly than your funding will allow. There might also be some response bias: Some students might be hesitant about criticizing the faculty (while others might be far too eager to do so).

3.130. (a) For the two factors (administration method, with three levels, and dosage, with two levels), the treatment combinations are shown in the table on the right, and the design is diagrammed below.

	Injection	Skin patch	IV drip
5 mg	1	2	3
10 mg	4	5	6

(b) Larger samples give more information; in particular, with large samples, we reduce the variability in the observed mean concentrations so that we can have more confidence that the differences we might observe are due to the treatment applied rather than random fluctuation.



3.131. Use a block design: Separate men and women, and randomly allocate each gender among the six treatments.

The remaining exercises relate to the material of Section 3.4 (Ethics). Answers are given for the first two; the rest call for student opinions, or information specific to the student's institution.

3.132. Parents who fail to return the consent form may be more likely to place less priority on education and therefore may give their children less help with homework, and so forth. Including those children in the control group is likely to lower that group's score.

Note: *This is a generalization, to be sure: We are not saying that* every *such parent does not value education, only that the percentage of this group that highly values education will almost certainly be lower than that percentage of the parents who return the form.*

3.133. The latter method (CASI) will show a higher percentage of drug use because respondents will generally be more comfortable (and more assured of anonymity) about revealing embarrassing or illegal behavior to a computer than to a person, so they will be more likely to be honest.

Chapter 4 Solutions

4.1. Only 6 of the first 20 digits on line 119 correspond to "heads," so the proportion of heads is $\frac{6}{20} = 0.3$. With such a small sample, random variation can produce results different from the expected value (0.5).

```
95857  07118  87664  92099
TTTTT  HTHHT  TTTTH  THHTT
```

4.2. The overall rate (56%) is an average. Graduation rates vary greatly among institutions; some will have have higher rates, and others lower.

4.3. **(a)** Most answers (99.5% of them) will be between 82% and 98%. **(b)** Based on 100,000 simulated trials—more than students are expected to do—the longest string of misses will be quite short (3 or fewer with probability 99%, 5 or fewer with probability 99.99%). The average ("expected") longest run of misses is about 1.7. For shots made, the average run is about 27, but there is lots of variation; 77% of simulations will have a longest run of made shots between 17 and 37, and about 95% of simulation will fall between 12 and 46.

4.5. If you hear music (or talking) one time, you will almost certainly hear the same thing for several more checks after that. (For example, if you tune in at the beginning of a 5-minute song and check back every 5 seconds, you'll hear that same song over 30 times.)

4.6. To estimate the probability, count the number of times the dice show 7 or 11, then divide by 25. For "perfectly made" (fair) dice, the number of winning rolls will nearly always (99.4% of the time) be between 1 and 11 out of 25.

4.7. Out of a very large number of patients taking this medication, the fraction who experience this bad side effect is about 0.00001.

Note: *Student explanations will vary, but should make clear that 0.00001 is a long-run average rate of occurrence. Because a probability of 0.00001 is often stated as "1 in in 100,000," it is tempting to interpret this probability as meaning "exactly 1 out of every 100,000." While we* expect *about 1 occurrence of side effects out of 100,000 patients, the actual number of side effects patients is random; it might be 0, or 1, or 2, ...*

4.8. **(a)** – **(c)** Results will vary, but after n tosses, the distribution of the proportion $\hat{p}$ is approximately Normal with mean 0.5 and standard deviation $0.5/\sqrt{n}$, while the distribution of the count of heads is approximately Normal with mean $0.5n$ and standard deviation $0.5\sqrt{n}$, so using the 68–95–99.7 rule, we have the results shown in the table on the right. For example, after 300 tosses, nearly all students should have $\hat{p}$ between 0.4134 and 0.5866, and a count between 124 and 176. Note that the range for $\hat{p}$ gets narrower, while the range for the count gets wider.

n	99.7% Range for $\hat{p}$	99.7% Range for count
50	0.5 ± 0.2121	25 ± 10.6
150	0.5 ± 0.1225	75 ± 18.4
300	0.5 ± 0.0866	150 ± 26.0
600	0.5 ± 0.0612	300 ± 36.7

4.9. The true probability (assuming perfectly fair dice) is $1 - \left(\frac{5}{6}\right)^4 \doteq 0.5177$, so students should conclude that the probability is "quite close to 0.5."

4.10. Sample spaces will likely include blonde, brunette (or brown), black, red, and gray. Depending on student imagination (and use of hair dye), other colors may be listed; there should at least be options to answer "other" and "bald."

4.11. One possibility: from 0 to __ hours (the largest number should be big enough to include all possible responses). In addition, some students might respond with fractional answers (e.g., 3.5 hours).

4.12. $P(\text{Black or White}) = 0.07 + 0.02 = 0.09$.

4.13. $P(\text{Blue, Green, Black, Brown, Grey, or White}) = 1 - P(\text{Purple, Red, Orange, or Yellow}) = 1 - (0.14 + 0.08 + 0.05 + 0.03) = 1 - 0.3 = 0.7$. Using Rule 4 (the complement rule) is slightly easier, because we only need to add the four probabilities of the colors we do *not* want, rather than adding the six probabilities of the colors we want.

4.14. $P(\text{not } 1) = 1 - 0.301 = 0.699$.

4.15. In Example 4.13, $P(B) = P(6 \text{ or greater})$ was found to be 0.222, so $P(A \text{ or } B) = P(A) + P(B) = 0.301 + 0.222 = 0.523$.

4.16. For each possible value (1, 2, . . . , 6), the probability is 1/6.

4.17. If T_k is the event "get tails on the kth flip," then T_1 and T_2 are independent, and $P(\text{two tails}) = P(T_1 \text{ and } T_2) = P(T_1)P(T_2) = \left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = \frac{1}{4}$.

4.18. If A_k is the event "the kth card drawn is an ace," then A_1 and A_2 are *not* independent; in particular, if we know that A_1 occurred, then the probability of A_2 is only $\frac{3}{51}$.

4.19. **(a)** The probability that both of two disjoint events occur is 0. (Multiplication is appropriate for *independent* events.) **(b)** Probabilities must be no more than 1; $P(A \text{ and } B)$ will be no more than 0.5. (We cannot determine this probability exactly from the given information.) **(c)** $P(A^c) = 1 - 0.35 = 0.65$.

4.20. **(a)** The two outcomes (say, A and B) in the sample space need not be equally likely. The only requirements are that $P(A) \geq 0$, $P(B) \geq 0$, and $P(A) + P(B) = 1$. **(b)** In a table of random digits, each digit has probability 0.1. **(c)** If A and B were independent, then $P(A \text{ and } B)$ would equal $P(A)P(B) = 0.06$. (That is, probabilities are multiplied, not added.) In fact, the given probabilities are impossible, because $P(A \text{ and } B)$ must be less than the smaller of $P(A)$ and $P(B)$.

4.21. There are six possible outcomes: { link1, link2, link3, link4, link5, leave }.

4.22. There are an infinite number of possible outcomes, and the description of the sample space will depend on whether the time is measured to any degree of accuracy (S is the set of all positive numbers) or rounded to (say) the nearest second ($S = \{0, 1, 2, 3, \ldots\}$), or nearest tenth of a second ($S = \{0, 0.1, 0.2, 0.3 \ldots\}$).

4.23. (a) P(“Empire State of Mind” or “I Gotta Feeling”) $= 0.180 + 0.068 = 0.248$.
(b) P(neither “Empire State of Mind” nor “I Gotta Feeling”) $= 1 - 0.248 = 0.752$.

4.24. (a) If R_k is the event “‘Party in the USA’ is the kth chosen ringtone,” then $P(R_1 \text{ and } R_2) = P(R_1)P(R_2) = 0.107^2 = 0.011449$. **(b)** The complement would be “at most one ringtone is ‘Party in the USA.’” $P[(R_1 \text{ and } R_2)^c] = 1 - P(R_1 \text{ and } R_2) = 0.988551$.

4.25. (a) The given probabilities have sum 0.97, so P(type AB) $= 0.03$.
(b) P(type O or B) $= 0.44 + 0.11 = 0.55$.

4.26. P(both are type O) $= (0.44)(0.52) = 0.2288$; P(both are the same type) $= (0.42)(0.35) + (0.11)(0.10) + (0.03)(0.03) + (0.44)(0.52) = 0.3877$.

4.27. (a) Not legitimate because the probabilities sum to 2. **(b)** Legitimate (for a nonstandard deck). **(c)** Legitimate (for a nonstandard die).

4.28. (a) The given probabilities have sum 0.77, so P(French) $= 0.23$.
(b) P(not English) $= 1 - 0.59 = 0.41$, using Rule 4. (Or, add the other three probabilities.)

4.29. (a) The given probabilities have sum 0.72, so this probability must be 0.28.
(b) P(at least a high school education) $= 1 - P$(has not finished HS) $= 1 - 0.12 = 0.88$. (Or add the other three probabilities.)

4.30. The probabilities of 2, 3, 4, and 5 are unchanged (1/6), so P(⚀ or ⚅) must still be 1/3. If P(⚅) $= 0.21$, then P(⚀) $= \frac{1}{3} - 0.21 = 0.12\overline{3}$ (or $\frac{37}{300}$). The complete table follows.

Face	⚀	⚁	⚂	⚃	⚄	⚅
Probability	$0.12\overline{3}$	1/6	1/6	1/6	1/6	0.21

4.31. For example, the probability for A-positive blood is $(0.42)(0.84) = 0.3528$ and for A-negative $(0.42)(0.16) = 0.0672$.

Blood type	A+	A–	B+	B–	AB+	AB–	O+	O–
Probability	0.3528	0.0672	0.0924	0.0176	0.0252	0.0048	0.3696	0.0704

4.32. (a) All are equally likely; the probability is 1/38. **(b)** Because 18 slots are red, the probability of a red is $P(\text{red}) = \frac{18}{38} \doteq 0.474$. **(c)** There are 12 winning slots, so P(win a column bet) $= \frac{12}{38} \doteq 0.316$.

4.33. **(a)** There are six arrangements of the digits 4, 9, and 1 (491, 419, 941, 914, 149, 194), so that $P(\text{win}) = \frac{6}{1000} = 0.006$. **(b)** The only winning arrangement is 222, so $P(\text{win}) = \frac{1}{1000} = 0.001$.

4.34. **(a)** There are $10^4 = 10{,}000$ possible PINs (0000 through 9999).* **(b)** The probability that a PIN has *no* 0s is 0.9^4 (because there are 9^4 PINs that can be made from the nine nonzero digits), so the probability of at least one 0 is $1 - 0.9^4 = 0.3439$.

*If we assume that PINs cannot have leading 0s, then there are only 9000 possible codes (1000–9999), and the probability of at least one 0 is $1 - \frac{9^4}{9000} = 0.271$.

4.35. $P(\text{none are O-negative}) = (1 - 0.07)^{10} \doteq 0.4840$, so $P(\text{at least one is O-negative}) \doteq 1 - 0.4840 = 0.5160$.

4.36. If we assume that each site is independent of the others (and that they can be considered as a random sample from the collection of sites referenced in scientific journals), then $P(\text{all seven are still good}) = 0.87^7 \doteq 0.3773$.

4.37. This computation would only be correct if the events "a randomly selected person is at least 75" and "a randomly selected person is a woman" were independent. This is likely not true; in particular, as women have a greater life expectancy than men, this fraction is probably greater than 3%.

4.38. As $P(\text{R}) = \frac{2}{6}$ and $P(\text{G}) = \frac{4}{6}$, and successive rolls are independent, the respective probabilities are:

$$\left(\tfrac{2}{6}\right)^4\left(\tfrac{4}{6}\right) = \tfrac{2}{243} \doteq 0.00823,\ \left(\tfrac{2}{6}\right)^4\left(\tfrac{4}{6}\right)^2 = \tfrac{4}{729} \doteq 0.00549, \text{ and } \left(\tfrac{2}{6}\right)^5\left(\tfrac{4}{6}\right) = \tfrac{2}{729} \doteq 0.00274.$$

4.39. **(a)** $(0.65)^3 \doteq 0.2746$ (under the random walk theory). **(b)** 0.35 (because performance in separate years is independent). **(c)** $(0.65)^2 + (0.35)^2 = 0.545$.

4.40. For any event A, along with its complement A^c, we have $P(S) = P(A \text{ or } A^c)$ because "A or A^c" includes all possible outcomes (that is, it is the entire sample space S). By Rule 2, $P(S) = 1$, and by Rule 3, $P(A \text{ or } A^c) = P(A) + P(A^c)$, because A and A^c are disjoint. Therefore, $P(A) + P(A^c) = 1$, from which Rule 4 follows.

4.41. Note that $A = (A \text{ and } B)$ or $(A \text{ and } B^c)$, and the events $(A \text{ and } B)$ and $(A \text{ and } B^c)$ are disjoint, so Rule 3 says that

$$P(A) = P\big((A \text{ and } B) \text{ or } (A \text{ and } B^c)\big) = P(A \text{ and } B) + P(A \text{ and } B^c).$$

If $P(A \text{ and } B) = P(A)P(B)$, then we have $P(A \text{ and } B^c) = P(A) - P(A)P(B) = P(A)(1 - P(B))$, which equals $P(A)P(B^c)$ by the complement rule.

4.42. **(a)** Hannah and Jacob's children can have alleles AA, BB, or AB, so they can have blood type A, B, or AB. (The table on the right shows the possible combinations.) **(b)** Either note that the four combinations in the table are equally likely, or compute:

	A	B
A	AA	AB
B	AB	BB

$$P(\text{type A}) = P(\text{A from Hannah and A from Jacob}) = P(\text{A}_H)P(\text{A}_J) = 0.5^2 = 0.25$$
$$P(\text{type B}) = P(\text{B from Hannah and B from Jacob}) = P(\text{B}_H)P(\text{B}_J) = 0.5^2 = 0.25$$
$$P(\text{type AB}) = P(\text{A}_H)P(\text{B}_J) + P(\text{B}_H)P(\text{A}_J) = 2 \cdot 0.25 = 0.5$$

4.43. **(a)** Nancy and David's children can have alleles BB, BO, or OO, so they can have blood type B or O. (The table on the right shows the possible combinations.) **(b)** Either note that the four combinations in the table are equally likely or compute $P(\text{type O}) = P(\text{O from Nancy and O from David}) = 0.5^2 = 0.25$ and $P(\text{type B}) = 1 - P(\text{type O}) = 0.75$.

	B	O
B	BB	BO
O	BO	OO

4.44. Any child of Jennifer and José has a 50% chance of being type A (alleles AA or AO), and each child inherits alleles independently of other children, so $P(\text{both are type A}) = 0.5^2 = 0.25$. For one child, we have $P(\text{type A}) = 0.5$ and $P(\text{type AB}) = P(\text{type B}) = 0.25$, so that $P(\text{both have same type}) = 0.5^2 + 0.25^2 + 0.25^2 = 0.375 = \frac{3}{8}$.

	A	O
A	AA	AO
B	AB	BO

4.45. **(a)** Any child of Jasmine and Joshua has an equal (1/4) chance of having blood type AB, A, B, or O (see the allele combinations in the table). Therefore, $P(\text{type O}) = 0.25$. **(b)** $P(\text{all three have type O}) = 0.25^3 = 0.015625 = \frac{1}{64}$. $P(\text{first has type O, next two do not}) = 0.25 \cdot 0.75^2 = 0.140625 = \frac{9}{64}$.

	A	O
B	AB	BO
O	AO	OO

4.46. $P(\text{grade of D or F}) = P(X = 0 \text{ or } X = 1) = 0.05 + 0.04 = 0.09$.

4.47. If H is the number of heads, then the distribution of H is as given on the right. $P(H = 0)$, the probability of two tails was previously computed in Exercise 4.17.

Value of H	0	1	2
Probabilities	1/4	1/2	1/4

4.48. $P(0.1 < X < 0.4) = 0.3$.

4.49. **(a)** The probabilities for a discrete *random variable* always add to one. **(b)** Continuous random variables can take values from any interval, not just 0 to 1. **(c)** A Normal random variable is continuous. (Also, a distribution is *associated with* a random variable, but "distribution" and "random variable" are not the same things.)

4.50. **(a)** If T is the event that a person uses Twitter, we can write the sample space as $\{T, T^c\}$. **(b)** There are various ways to express this; one would be

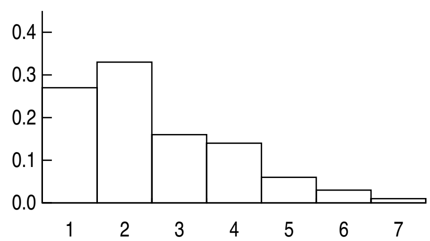
$$\{TTT, TTT^c, TT^cT, T^cTT, TT^cT^c, T^cTT^c, T^cT^cT, T^cT^cT^c\}.$$

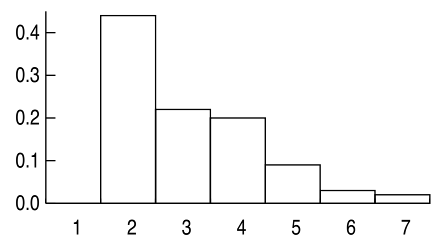
(c) For this random variable (call it X), the sample space is $\{0, 1, 2, 3\}$. **(d)** The sample space in part (b) reveals which of the three people use Twitter. This may or may not be important information; it depends on what questions we wish to ask about our sample.

4.51. **(a)** Based on the information from Exercise 4.50, along with the complement rule, $P(T) = 0.19$ and $P(T^c) = 0.81$. **(b)** Use the multiplication rule for independent events; for example, $P(TTT) = 0.19^3 \doteq 0.0069$, $P(TTT^c) = (0.19^2)(0.81) \doteq 0.0292$, $P(TT^cT^c) = (0.19)(0.81^2) \doteq 0.1247$, and $P(T^cT^cT^c) = 0.81^3 \doteq 0.5314$. **(c)** Add up the probabilities from (b) that correspond to each value of X.

Outcome	TTT	TTT^c	TT^cT	T^cTT	T^cT^cT	T^cTT^c	TT^cT^c	$T^cT^cT^c$
Probability	0.0069	0.0292	0.0292	0.0292	0.1247	0.1247	0.1247	0.5314
Value of X	0		1			2		3
Probability	0.0069		0.0877			0.3740		0.5314

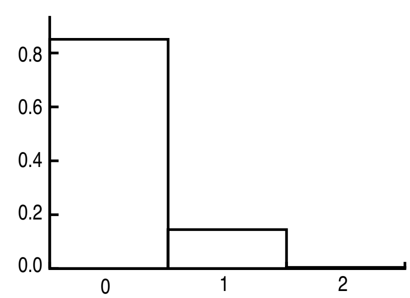
4.52. The two histograms are shown below. The most obvious difference is that a "family" must have at least two people. Otherwise, the family-size distribution has slightly larger probabilities for 2, 3, or 4, while for large family/household sizes, the differences between the distributions are small.





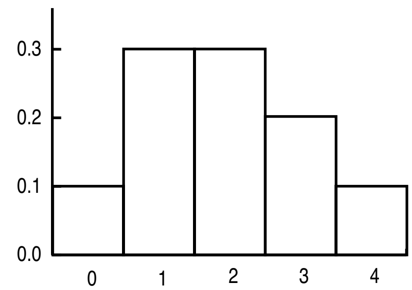
4.53. **(a)** See also the solution to Exercise 4.22. If we view this time as being measured to any degree of accuracy, it is continuous; if it is rounded, it is discrete. **(b)** A count like this must be a whole number, so it is discrete. **(c)** Incomes—whether given in dollars and cents, or rounded to the nearest dollar—are discrete. (However, it is often useful to treat such variables as continuous.)

4.54. **(a)** $0.8507 + 0.1448 + 0.0045 = 1$. **(b)** Histogram on the right. (The third bar is so short that it blends in with the horizontal axis.) **(c)** $P(\text{at least one ace}) = 0.1493$, which can be computed either as $0.1448 + 0.0045$ or $1 - 0.8507$.

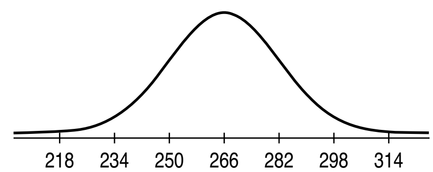


4.55. **(a)** Histogram on the right. **(b)** "At least one nonword error" is the event "$X \geq 1$" (or "$X > 0$"). $P(X \geq 1) = 1 - P(X = 0) = 0.9$. **(c)** "$X \leq 2$" is "no more than two nonword errors," or "fewer than three nonword errors."

$$P(X \leq 2) = 0.7 = P(X = 0) + P(X = 1) + P(X = 2)$$
$$= 0.1 + 0.3 + 0.3$$
$$P(X < 2) = 0.4 = P(X = 0) + P(X = 1) = 0.1 + 0.3$$

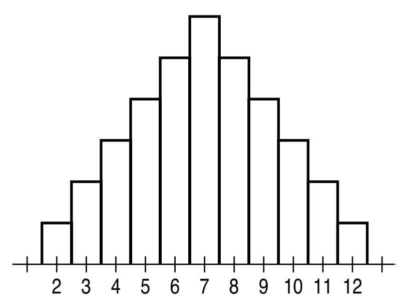


4.56. **(a)** Curve on the right. A good procedure is to draw the curve first, locate the points where the curvature changes, then mark the horizontal axis. Students may at first make mistakes like drawing a half-circle instead of the correct "bell-shaped" curve or being careless about locating the standard deviation. **(b)** About 0.81: $P(Y \leq 280) = P\left(\frac{Y-266}{16} \leq \frac{280-266}{16}\right) = P(Z \leq 0.875)$. Software gives 0.8092; Table A gives 0.8078 for 0.87 and 0.8106 for 0.88 (so the average is again 0.8092).



4.57. **(a)** The pairs are given below. We must assume that we can distinguish between, for example, "(1,2)" and "(2,1)"; otherwise, the outcomes are not equally likely. **(b)** Each pair has probability 1/36. **(c)** The value of X is given below each pair. For the distribution (given below), we see (for example) that there are four pairs that add to 5, so $P(X = 5) = \frac{4}{36}$. Histogram below, right. **(d)** $P(7 \text{ or } 11) = \frac{6}{36} + \frac{2}{36} = \frac{8}{36} = \frac{2}{9}$. **(e)** $P(\text{not } 7) = 1 - \frac{6}{36} = \frac{5}{6}$.

(1,1) 2	(1,2) 3	(1,3) 4	(1,4) 5	(1,5) 6	(1,6) 7
(2,1) 3	(2,2) 4	(2,3) 5	(2,4) 6	(2,5) 7	(2,6) 8
(3,1) 4	(3,2) 5	(3,3) 6	(3,4) 7	(3,5) 8	(3,6) 9
(4,1) 5	(4,2) 6	(4,3) 7	(4,4) 8	(4,5) 9	(4,6) 10
(5,1) 6	(5,2) 7	(5,3) 8	(5,4) 9	(5,5) 10	(5,6) 11
(6,1) 7	(6,2) 8	(6,3) 9	(6,4) 10	(6,5) 11	(6,6) 12



Value of X	2	3	4	5	6	7	8	9	10	11	12
Probability	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

4.58. The possible values of Y are 1, 2, 3, . . . , 12, each with probability 1/12. Aside from drawing a diagram showing all the possible combinations, one can reason that the first (regular) die is equally likely to show any number from 1 through 6. Half of the time, the second roll shows 0, and the rest of the time it shows 6. Each possible outcome therefore has probability $\frac{1}{6} \cdot \frac{1}{2}$.

4.59. The table on the right shows the additional columns to add to the table given in the solution to Exercise 4.57. There are 48 possible (equally-likely) combinations.

(1,7) 8	(1,8) 9
(2,7) 9	(2,8) 10
(3,7) 10	(3,8) 11
(4,7) 11	(4,8) 12
(5,7) 12	(5,8) 13
(6,7) 13	(6,8) 14

Value of X	2	3	4	5	6	7	8	9	10	11	12	13	14
Probability	$\frac{1}{48}$	$\frac{2}{48}$	$\frac{3}{48}$	$\frac{4}{48}$	$\frac{5}{48}$	$\frac{6}{48}$	$\frac{6}{48}$	$\frac{6}{48}$	$\frac{5}{48}$	$\frac{4}{48}$	$\frac{3}{48}$	$\frac{2}{48}$	$\frac{1}{48}$

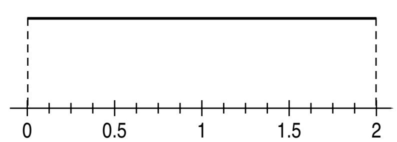
4.60. (a) W can be 0, 1, 2, or 3. **(b)** See the top two lines of the table below. **(c)** The distribution is given in the bottom two lines of the table. For example, $P(W = 0) = (0.73)(0.73)(0.73) \doteq 0.3890$, and in the same way, $P(W = 3) = 0.27^3 \doteq 0.1597$. For $P(W = 1)$, note that each of the three arrangements that give $W = 1$ have probability $(0.73)(0.73)(0.27) = 0.143883$, so $P(W = 1) = 3(0.143883) \doteq 0.4316$. Similarly, $P(W = 2) = 3(0.73)(0.27)(0.27) \doteq 0.1597$.

Arrangement	DDD	DDF	DFD	FDD	FFD	FDF	DFF	FFF
Probability	0.3890	0.1439	0.1439	0.1439	0.0532	0.0532	0.0532	0.0197
Value of W	0		1			2		3
Probability	0.3890		0.4316			0.1597		0.0197

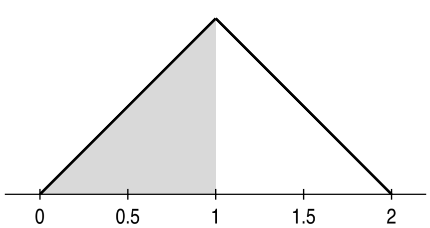
4.61. (a) $P(X < 0.6) = 0.6$. **(b)** $P(X \leq 0.6) = 0.6$. **(c)** For continuous random variables, "equal to" has no effect on the probability; that is, $P(X = c) = 0$ for any value of c.

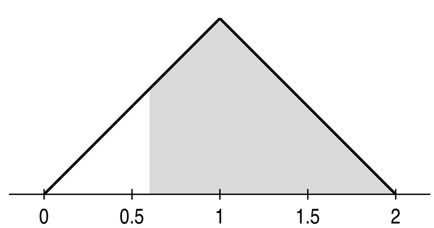
4.62. (a) $P(X \geq 0.30) = 0.7$. **(b)** $P(X = 0.30) = 0$. **(c)** $P(0.30 < X < 1.30) = P(0.30 < X < 1) = 0.7$. **(d)** $P(0.20 \leq X \leq 0.25 \text{ or } 0.7 \leq X \leq 0.9) = 0.05 + 0.2 = 0.25$. **(e)** $P(\text{not } [0.4 \leq X \leq 0.7]) = 1 - P(0.4 \leq X \leq 0.7) = 1 - 0.3 = 0.7$.

4.63. (a) The height should be $\frac{1}{2}$ since the area under the curve must be 1. The density curve is at the right. **(b)** $P(Y \leq 1.6) = \frac{1.6}{2} = 0.8$. **(c)** $P(0.5 < Y < 1.7) = \frac{1.2}{2} = 0.6$. **(d)** $P(Y \geq 0.95) = \frac{1.05}{2} = 0.525$.



4.64. (a) The area of a triangle is $\frac{1}{2}bh = \frac{1}{2}(2)(1) = 1$. **(b)** $P(Y < 1) = 0.5$. **(c)** $P(Y > 0.6) = 0.82$; the easiest way to compute this is to note that the *unshaded* area is a triangle with area $\frac{1}{2}(0.6)(0.6) = 0.18$.





4.65. $P(8 \le \bar{x} \le 10) = P\left(\frac{8-9}{0.0724} \le \frac{\bar{x}-9}{0.0724} \le \frac{10-9}{0.0724}\right) = P(-13.8 \le Z \le 13.8)$. This probability is essentially 1; $\bar{x}$ will almost certainly estimate μ within ± 1 (in fact, it will almost certainly be much closer than this).

4.66. (a) $P(0.52 \le \hat{p} \le 0.60) = P\left(\frac{0.52-0.56}{0.019} \le \frac{\hat{p}-0.56}{0.019} \le \frac{0.60-0.56}{0.019}\right) = P(-2.11 \le Z \le 2.11) = 0.9826 - 0.0174 = 0.9652$. **(b)** $P(\hat{p} \ge 0.72) = P\left(\frac{\hat{p}-0.56}{0.019} \ge \frac{0.72-0.56}{0.019}\right) = P(Z \ge 8.42)$; this is basically 0.

4.67. The possible values of X are \$0 and \$1, each with probability 0.5 (because the coin is fair). The mean is $\$0\left(\frac{1}{2}\right) + \$1\left(\frac{1}{2}\right) = \$0.50$.

4.69. If $Y = 15 + 8X$, then $\mu_Y = 15 + 8\mu_X = 15 + 8(10) = 95$.

4.70. If $W = 0.5U + 0.5V$, then $\mu_W = 0.5\mu_U + 0.5\mu_V = 0.5(20) + 0.5(20) = 20$.

4.71. First we note that $\mu_X = 0(0.5) + 2(0.5) = 1$, so $\sigma_X^2 = (0-1)^2(0.5) + (2-1)^2(0.5) = 1$ and $\sigma_X = \sqrt{\sigma_X^2} = 1$.

4.72. (a) Each toss of the coin is independent (that is, coins have no memory). **(b)** The variance is multiplied by $10^2 = 100$. (The mean and *standard deviation* are multiplied by 10.) **(c)** The correlation does not affect the mean of a sum (although it does affect the variance and standard deviation).

4.73. The mean is

$$\mu_X = (0)(0.3) + (1)(0.1) + (2)(0.1) + (3)(0.2) + (4)(0.1) + (5)(0.2) = 2.3 \text{ servings.}$$

The variance is

$$\begin{aligned}\sigma_X^2 = &(0-2.3)^2(0.3) + (1-2.3)^2(0.1) + (2-2.3)^2(0.1)\\ &+ (3-2.3)^2(0.2) + (4-2.3)^2(0.1) + (5-2.3)^2(0.2) = 3.61,\end{aligned}$$

so the standard deviation is $\sigma_X = \sqrt{3.61} = 1.9$ servings.

4.74. The mean number of aces is $\mu_X = (0)(0.8507) + (1)(0.1448) + (2)(0.0045) = 0.1538$.

Note: *The exact value of the mean is 2/13, because 1/13 of the cards are aces, and two cards have been dealt to us.*

4.75. The average grade is $\mu = (0)(0.05)+(1)(0.04)+(2)(0.20)+(3)(0.40)+(4)(0.31) = 2.88$.

4.76. The means are

$$\begin{aligned}(0)(0.1) + (1)(0.3) + (2)(0.3) + (3)(0.2) + (4)(0.1) &= 1.9 \text{ nonword errors and}\\ (0)(0.4) + (1)(0.3) + (2)(0.2) + (3)(0.1) &= 1 \text{ word error}\end{aligned}$$

4.77. In the solution to Exercise 4.74, we found $\mu_X = 0.1538$ aces, so

$$\sigma_X^2 = (0 - 0.1538)^2(0.8507) + (1 - 0.1538)^2(0.1448) + (2 - 0.1538)^2(0.0045) \doteq 0.1391,$$

and the standard deviation is $\sigma_X \doteq \sqrt{0.1391} \doteq 0.3730$ aces.

4.78. In the solution to Exercise 4.75, we found the average grade was $\mu = 2.88$, so

$$\sigma^2 = (0 - 2.88)^2(0.05) + (1 - 2.88)^2(0.04) + (2 - 2.88)^2(0.2) + (3 - 2.88)^2(0.4) + (4 - 2.88)^2(0.31) = 1.1056,$$

and the standard deviation is $\sigma = \sqrt{1.1056} \doteq 1.0515$.

4.79. **(a)** With $\rho = 0$, the variance is $\sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2 = (75)^2 + (41)^2 = 7306$, so the standard deviation is $\sigma_{X+Y} = \sqrt{7306} \doteq \$85.48$. **(b)** This is larger; the negative correlation decreased the variance.

4.80. **(a)** The mean of Y is $\mu_Y = 1$—the obvious balance point of the triangle. **(b)** Both X_1 and X_2 have mean $\mu_{X_1} = \mu_{X_2} = 0.5$ and $\mu_Y = \mu_{X_1} + \mu_{X_2}$.

4.81. The situation described in this exercise—"people who have high intakes of calcium in their diets are more compliant than those who have low intakes"—implies a positive correlation between calcium intake and compliance. Because of this, the variance of total calcium intake is greater than the variance we would see if there were no correlation (as the calculations in Example 4.38 demonstrate).

4.82. Let N and W be nonword and word error counts. In Exercise 4.76, we found $\mu_N = 1.9$ errors and $\mu_W = 1$ error. The variances of these distributions are $\sigma_N^2 = 1.29$ and $\sigma_W^2 = 1$, so the standard deviations are $\sigma_N \doteq 1.1358$ errors and $\sigma_W = 1$ error. The mean total error count is $\mu_N + \mu_W = 2.9$ errors for both cases. **(a)** If error counts are independent (so that $\rho = 0$), $\sigma_{N+W}^2 = \sigma_N^2 + \sigma_W^2 = 2.29$ and $\sigma_{N+W} \doteq 1.5133$ errors. (Note that we add the *variances,* not the standard deviations.) **(b)** With $\rho = 0.5$, $\sigma_{N+W}^2 = \sigma_N^2 + \sigma_W^2 + 2\rho\sigma_N\sigma_W \doteq 2.29 + 1.1358 = 3.4258$ and $\sigma_{N+W} \doteq 1.8509$ errors.

4.83. **(a)** The mean for one coin is $\mu_1 = (0)\left(\frac{1}{2}\right) + (1)\left(\frac{1}{2}\right) = 0.5$ and the variance is $\sigma_1^2 = (0 - 0.5)^2\left(\frac{1}{2}\right) + (1 - 0.5)^2\left(\frac{1}{2}\right) = 0.25$, so the standard deviation is $\sigma_1 = 0.5$. **(b)** Multiply μ_1 and σ_1^2 by 4: $\mu_4 = 4\mu_1 = 2$ and $\sigma_4^2 = 4\sigma_1^4 = 1$, so $\sigma_4 = 1$. **(c)** Note that because of the symmetry of the distribution, we do not need to compute the mean to see that $\mu_4 = 2$; this is the obvious balance point of the probability histogram in Figure 4.7. The details of the two computations are

$$\mu_W = (0)(0.0625) + (1)(0.25) + (2)(0.375) + (3)(0.25) + (4)(0.0625) = 2$$

$$\sigma_W^2 = (0 - 2)^2(0.0625) + (1 - 2)^2(0.25) + (2 - 2)^2(0.375) + (3 - 2)^2(0.25) + (4 - 2)^2(0.0625) = 1.$$

4.84. If D is the result of rolling a single four-sided die, then $\mu_D = (1 + 2 + 3 + 4)\left(\frac{1}{4}\right) = 2.5$, and $\sigma_D^2 = [(1 - 2.5)^2 + (2 - 2.5)^2 + (3 - 2.5)^2 + (4 - 2.5)^2]\frac{1}{4} = 1.25$. Then for the sum

$I = D_1 + D_2 + 1$, we have mean intelligence $\mu_I = 2\mu_D + 1 = 6$. The variance of I is $\sigma_I^2 = 2\sigma_D^2 = 2.5$, so $\sigma_I \doteq 1.5811$.

4.85. With R as the rod length and B_1 and B_2 the bearing lengths, we have $\mu_{B_1+R+B_2} = 12 + 2 \cdot 2 = 16$ cm and $\sigma_{B_1+R+B_2} = \sqrt{0.004^2 + 2 \cdot 0.001^2} \doteq 0.004243$ mm.

4.86. **(a)** $d_1 = 2\sigma_R = 0.008$ mm $= 0.0008$ cm and $d_2 = 2\sigma_B = 0.002$ mm $= 0.0002$ cm. **(b)** The natural tolerance of the assembled parts is $2\sigma_{B_1+R+B_2} \doteq 0.008485$ mm $=$ 0.0008485 cm.

4.87. **(a)** Not independent: Knowing the total X of the first two cards tells us something about the total Y for three cards. **(b)** Independent: Separate rolls of the dice should be independent.

4.88. Divide the given values by 2.54: $\mu \doteq 69.6063$ in and $\sigma \doteq 2.8346$ in.

4.89. With $\rho = 1$, we have:

$$\sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2 + 2\rho\sigma_X\sigma_Y = \sigma_X^2 + \sigma_Y^2 + 2\sigma_X\sigma_Y = (\sigma_X + \sigma_Y)^2$$

And of course, $\sigma_{X+Y} = \sqrt{(\sigma_X + \sigma_Y)^2} = \sigma_X + \sigma_Y$.

4.90. The mean of X is $(\mu - \sigma)(0.5) + (\mu + \sigma)(0.5) = \mu$, and the standard deviation is $\sqrt{(\mu - \sigma - \mu)^2(0.5) + (\mu + \sigma - \mu)^2(0.5)} = \sqrt{\sigma^2} = \sigma$.

4.91. Although the probability of having to pay for a total loss for one or more of the 10 policies is very small, if this were to happen, it would be financially disastrous. On the other hand, for thousands of policies, the law of large numbers says that the average claim on many policies will be close to the mean, so the insurance company can be assured that the premiums they collect will (almost certainly) cover the claims.

4.92. The total loss T for 10 fires has mean $\mu_T = 10 \cdot \$300 = \$3000$, and standard deviation $\sigma_T = \sqrt{10 \cdot \$400^2} = \$400\sqrt{10} \doteq \$1264.91$. The average loss is $T/10$, so $\mu_{T/10} = \frac{1}{10}\mu_T = \$300$ and $\sigma_{T/10} = \frac{1}{10}\sigma_T \doteq \$126.49$.

The total loss T for 12 fires has mean $\mu_T = 12 \cdot \$300 = \$3600$, and standard deviation $\sigma_T = \sqrt{12 \cdot \$400^2} = \$400\sqrt{12} \doteq \$1385.64$. The average loss is $T/12$, so $\mu_{T/12} = \frac{1}{12}\mu_T = \$300$ and $\sigma_{T/12} = \frac{1}{12}\sigma_T \doteq \$115.47$.

Note: *The mean of the average loss is the same regardless of the number of policies, but the standard deviation decreases as the number of policies increases. With thousands of policies, the standard deviation is very small, so the average claim will be close to $300, as was stated in the solution to the previous problem.*

4.93. **(a)** Add up the given probabilities and subtract from 1; this gives $P(\text{man does not die in the next five years}) = 0.99749$. **(b)** The distribution of income (or loss) is given below. Multiplying each possible value by its probability gives the mean intake $\mu \doteq \$623.22$.

Age at death	21	22	23	24	25	Survives
Loss or income	−\$99,825	−\$99,650	−\$99,475	−\$99,300	−\$99,125	\$875
Probability	0.00039	0.00044	0.00051	0.00057	0.00060	0.99749

4.94. The mean μ of the company's "winnings" (premiums) and their "losses" (insurance claims) is positive. Even though the company will lose a large amount of money on a small number of policyholders who die, it will gain a small amount on the majority. The law of large numbers says that the average "winnings" minus "losses" should be close to μ, and overall the company will almost certainly show a profit.

4.95. The events "roll a 3" and "roll a 5" are disjoint, so $P(3 \text{ or } 5) = P(3) + P(5) = \frac{1}{6} + \frac{1}{6} = \frac{2}{6} = \frac{1}{3}$.

4.96. The events E (roll is even) and G (roll is greater than 4) are *not* disjoint—specifically, $E \text{ and } G = \{6\}$—so $P(E \text{ or } G) = P(E) + P(G) - P(E \text{ and } G) = \frac{3}{6} + \frac{2}{6} - \frac{1}{6} = \frac{4}{6} = \frac{2}{3}$.

4.97. Let A be the event "next card is an ace" and B be "two of Slim's four cards are aces." Then, $P(A \mid B) = \frac{2}{48}$ because (other than those in Slim's hand) there are 48 cards, of which 2 are aces.

4.98. Let A_1 = "the next card is a diamond" and A_2 = "the second card is a diamond." We wish to find $P(A_1 \text{ and } A_2)$. There are 27 unseen cards, of which 10 are diamonds, so $P(A_1) = \frac{10}{27}$, and $P(A_2 \mid A_1) = \frac{9}{26}$, so $P(A_1 \text{ and } A_2) = \frac{10}{27} \times \frac{9}{26} = \frac{5}{39} \doteq 0.1282$.

Note: *Technically, we wish to find* $P(A_1 \text{ and } A_2 \mid B)$*, where B is the given event (25 cards visible, with 3 diamonds in Slim's hand). We have* $P(A_1 \mid B) = \frac{10}{27}$ *and* $P(A_2 \mid A_1 \text{ and } B) = \frac{9}{26}$*, and compute* $P(A_1 \text{ and } A_2 \mid B) = P(A_1 \mid B) \times P(A_2 \mid A_1 \text{ and } B)$.

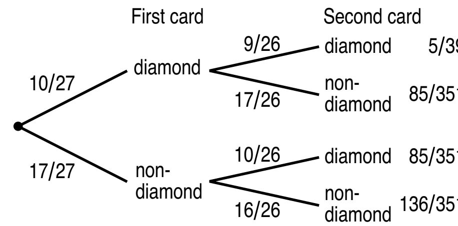
4.99. This computation uses the addition rule for disjoint events, which is appropriate for this setting because B (full-time students) is made up of four disjoint groups (those in each of the four age groups).

4.100. With A and B as defined in Example 4.44 (respectively, 15- to 19-year-old students, and full-time students), we want to find

$$P(B \mid A) = \frac{P(B \text{ and } A)}{P(A)} = \frac{0.21}{0.21 + 0.02} \doteq 0.9130$$

For these two calculations, we restrict our attention to different subpopulations of students (that is, different rows of the table given in Example 4.44). For $P(A \mid B)$, we ask what fraction of full-time students (the subpopulation) are aged 15 to 19 years. For $P(B \mid A)$, we ask what fraction of the subpopulation of 15- to 19-year-old students are full-time.

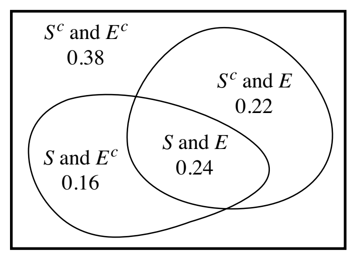
4.101. The tree diagram shows the probability found in Exercise 4.98 on the top branch. The middle two branches (added together) give the probability that Slim gets exactly one diamond from the next two cards, and the bottom branch is the probability that neither card is a diamond.



4.102. **(a)** The given statement is only true for disjoint events; in general, $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$. **(b)** $P(A)$ *plus* $P(A^c)$ is always equal to one. **(c)** Two events are *independent* if $P(B \mid A) = P(B)$. They are *disjoint* if $P(A \text{ and } B) = 0$.

4.103. For a randomly chosen adult, let S = "(s)he gets enough sleep" and let E = "(s)he gets enough exercise," so $P(S) = 0.4$, $P(E) = 0.46$, and $P(S \text{ and } E) = 0.24$. **(a)** $P(S \text{ and } E^c) = 0.4 - 0.24 = 0.16$. **(b)** $P(S^c \text{ and } E) = 0.46 - 0.24 = 0.22$. **(c)** $P(S^c \text{ and } E^c) = 1 - (0.4 + 0.46 - 0.24) = 0.38$. **(d)** The answers in (a) and (b) are found by a variation of the addition rule for disjoint events: We note that $P(S) = P(S \text{ and } E) + P(S \text{ and } E^c)$ and $P(E) = P(S \text{ and } E) + P(S^c \text{ and } E)$. In each case, we know the first two probabilities, and find the third by subtraction. The answer for (c) is found by using the general addition rule to find $P(S \text{ or } E)$, and noting that $S^c \text{ and } E^c = (S \text{ or } E)^c$.

4.104. With S and E as defined in the previous solution, the Venn diagram on the right illustrates the probabilities computed above.



4.105. For a randomly chosen high school student, let L = "student admits to lying" and M = "student is male," so $P(L) = 0.48$, $P(M) = 0.5$, and $P(M \text{ and } L) = 0.25$. Then $P(M \text{ or } L) = P(M) + P(L) - P(M \text{ and } L) = 0.73$.

4.106. Using the addition rule for disjoint events, note that $P(M^c \text{ and } L) = P(L) - P(M \text{ and } L) = 0.23$. Then by the definition of conditional probability, $P(M^c \mid L) = \dfrac{P(M \text{ and } L)}{P(L)} = \dfrac{0.23}{0.48} \doteq 0.4792$.

4.107. Let B = "student is a binge drinker" and M = "student is male." **(a)** The four probabilities sum to $0.11 + 0.12 + 0.32 + 0.45 = 1$. **(b)** $P(B^c) = 0.32 + 0.45 = 0.77$. **(c)** $P(B^c \mid M) = \frac{P(B^c \text{ and } M)}{P(M)} = \frac{0.32}{0.11 + 0.32} \doteq 0.7442$. **(d)** In the language of this chapter, the events are not independent. An attempt to phrase this for someone who has not studied this material might say something like, "Knowing a student's gender gives some information about whether or not that student is a binge drinker."

Note: *Specifically, male students are slightly more likely to be binge drinkers. This statement might surprise students who look at the table and note that the proportion of binge drinkers in the men's columns is smaller than that proportion in the women's* *column. We cannot compare those proportions directly; we need to compare the* conditional *probabilities of binge drinkers within each given gender (see the solution to the next exercise.)*

4.108. Let B = "student is a binge drinker" and M = "student is male." **(a)** These two probabilities are given as entries in the table: $P(M \text{ and } B) = 0.11$ and $P(M^c \text{ and } B) = 0.12$. **(b)** These are conditional probabilities: $P(B \mid M) = \frac{P(B \text{ and } M)}{P(M)} = \frac{0.11}{0.11+0.32} \doteq 0.2558$ and $P(B \mid M^c) = \frac{P(B \text{ and } M^c)}{P(M^c)} = \frac{0.12}{0.12+0.45} \doteq 0.2105$. **(c)** The fact that $P(B \mid M) > P(B \mid M^c)$ indicates that male students are more likely to be binge drinkers (see the comment in the solution to the previous exercise). The other comparison, $P(M \text{ and } B) < P(M^c \text{ and } B)$, is more a reflection of the fact that the survey reported responses for more women (57%) than men (43%) and does not by itself allow for comparison of binge drinking between the genders. (To understand this better, imagine a more extreme case, where, say, 90% of respondents were women ....)

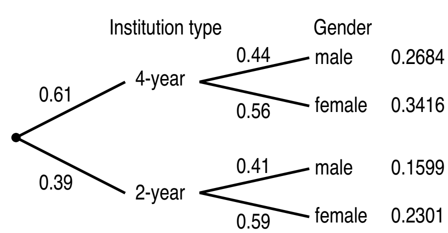
4.109. Let M = "male" and C = "attends a 4-year institution." (C is not an obvious choice, but it is less confusing than F, which we might mistake for "female.") We have been given $P(C) = 0.61$, $P(C^c) = 0.39$, $P(M \mid C) = 0.44$ and $P(M \mid C^c) = 0.41$. **(a)** To create the table, observe that:

	Men	Women
4-year	0.2684	0.3416
2-year	0.1599	0.2301

$$P(M \text{ and } C) = P(M \mid C)P(C) = (0.44)(0.61) = 0.2684$$

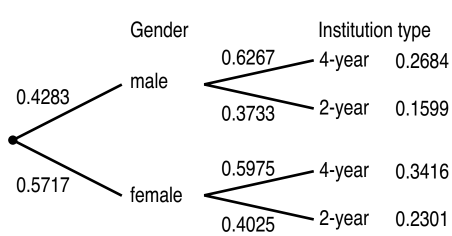
And similarly, $P(M \text{ and } C^c) = P(M \mid C^c)P(C^c) = (0.41)(0.39) = 0.1599$. The other two entries can be found in a similar fashion or by observing that, for example, the two numbers on the first row must sum to $P(C) = 0.61$.
(b) $P(C \mid M^c) = \frac{P(C \text{ and } M^c)}{P(M^c)} = \frac{0.3416}{0.3416+0.2301} \doteq 0.5975$.

4.110. The branches of this tree diagram have the probabilities given in Exercise 4.109, and the branches end with the probabilities found in the solution to that exercise.



4.111. As before, let M = "male" and C = "attends a 4-year institution." For this tree diagram, we need to compute $P(M) = 0.2684 + 0.1599 = 0.4283$, $P(M^c) = 0.3416 + 0.2301 = 0.5717$, as well as $P(C \mid M)$, $P(C \mid M^c)$, $P(C^c \mid M)$, and $P(C^c \mid M^c)$. For example,

$$P(C \mid M) = \frac{P(C \text{ and } M)}{P(M)} = \frac{0.2684}{0.4283} \doteq 0.6267.$$

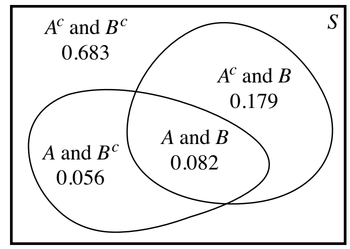


All the computations for this diagram are "inconvenient" because they require that we work *backward* from the ending probabilities, instead of working *forward* from the given probabilities (as we did in the previous tree diagram).

4.112. $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B) = 0.138 + 0.261 - 0.082 = 0.317$.

4.113. $P(A \mid B) = \frac{P(A \text{ and } B)}{P(B)} = \frac{0.082}{0.261} \doteq 0.3142$. If A and B were independent, then $P(A \mid B)$ would equal $P(A)$, and also $P(A \text{ and } B)$ would equal the product $P(A)P(B)$.

4.114. (a) $\{A \text{ and } B\}$ means the selected household is both prosperous and educated. $P(A \text{ and } B) = 0.082$ (given).
(b) $\{A \text{ and } B^c\}$ means the household is prosperous but not educated. $P(A \text{ and } B^c) = P(A) - P(A \text{ and } B) = 0.056$.
(c) $\{A^c \text{ and } B\}$ means the household is not prosperous but is educated. $P(A^c \text{ and } B) = P(B) - P(A \text{ and } B) = 0.179$.
(d) $\{A^c \text{ and } B^c\}$ means the household is neither prosperous nor educated. $P(A^c \text{ and } B^c) = 0.683$ (so that the probabilities add to 1).



4.115. (a) "The vehicle is a light truck" $= A^c$; $P(A^c) = 0.69$.
(b) "The vehicle is an imported car" $= A \text{ and } B$. To find this probability, note that we have been given $P(B^c) = 0.78$ and $P(A^c \text{ and } B^c) = 0.55$. From this we can determine that 78% − 55% = 23% of vehicles sold were domestic cars—that is, $P(A \text{ and } B^c) = 0.23$—so $P(A \text{ and } B) = P(A) - P(A \text{ and } B^c) = 0.31 - 0.23 = 0.08$.

	$P(A) = 0.31$	$P(A^c) = \mathbf{0.69}$
$P(B) = 0.22$	$P(A \text{ and } B) = 0.08$	$P(A^c \text{ and } B) = 0.14$
$P(B^c) = \mathbf{0.78}$	$P(A \text{ and } B^c) = 0.23$	$P(A^c \text{ and } B^c) = \mathbf{0.55}$

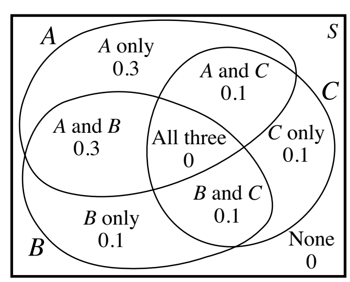
Note: *The table shown here summarizes all that we can determine from the given information* (**bold**).

4.116. Let A be the event "income $\geq$ \$1 million" and B be "income $\geq$ \$100,000." Then "A and B" is the same as A, so:

$$P(A \mid B) = \frac{P(A)}{P(B)} = \frac{\frac{392{,}220}{142{,}978{,}806}}{\frac{17{,}993{,}498}{142{,}978{,}806}} = \frac{392{,}220}{17{,}993{,}498} \doteq 0.02180$$

4.117. See also the solution to Exercise 4.115, especially the table of probabilities given there. **(a)** $P(A^c \mid B) = \frac{P(A^c \text{ and } B)}{P(B)} = \frac{0.14}{0.22} \doteq 0.6364$. **(b)** The events A^c and B are *not* independent; if they were, $P(A^c \mid B)$ would be the same as $P(A^c) = 0.69$.

4.118. To find the probabilities in this Venn diagram, begin with $P(A \text{ and } B \text{ and } C) = 0$ in the center of the diagram. Then each of the two-way intersections $P(A \text{ and } B)$, $P(A \text{ and } C)$, and $P(B \text{ and } C)$ go in the remainder of the overlapping areas; if $P(A \text{ and } B \text{ and } C)$ had been something other than 0, we would have subtracted this from each of the two-way intersection probabilities to find, for example, $P(A \text{ and } B \text{ and } C^c)$. Next, determine $P(A \text{ only})$ so that the total probability of the regions that make up the event A is 0.7. Finally, $P(\text{none}) = P(A^c \text{ and } B^c \text{ and } C^c) = 0$ because the total probability inside the three sets A, B, and C is 1.



4.119. We seek $P(\text{at least one offer}) = P(A \text{ or } B \text{ or } C)$; we can find this as $1 - P(\text{no offers}) = 1 - P(A^c \text{ and } B^c \text{ and } C^c)$. We see in the Venn diagram of Exercise 4.118 that this probability is 1.

4.120. This is $P(A \text{ and } B \text{ and } C^c)$. As was noted in Exercise 4.118, because $P(A \text{ and } B \text{ and } C) = 0$, this is the same as $P(A \text{ and } B) = 0.3$.

4.121. $P(B \mid C) = \dfrac{P(B \text{ and } C)}{P(C)} = \frac{0.1}{0.3} = \frac{1}{3}$. $P(C \mid B) = \dfrac{P(B \text{ and } C)}{P(B)} = \frac{0.1}{0.5} = 0.2$.

4.122. Let W = "the degree was earned by a woman" and P = "the degree was a professional degree." **(a)** To construct the table (below), divide each entry by the grand total of all entries (2403); for example, $\frac{933}{2403} \doteq 0.3883$ is the fraction of all degrees that were bachelor's degrees awarded to women. Some students may also find the row totals (1412 and 991) and the column totals (1594, 662, 95, 52) and divide those by the grand total; for example, $\frac{1594}{2403} \doteq 0.6633$ is the fraction of all degrees that were bachelor's degrees. **(b)** $P(W) = \frac{1412}{2403} \doteq 0.5876$ (this is one of the optional marginal probabilities from the table below). **(c)** $P(W \mid P) = \frac{51/2403}{95/2403} = \frac{51}{95} \doteq 0.5368$. (This is the "Female" entry from the "Professional" column, divided by that column's total.) **(d)** W and P are *not* independent; if they were, the two probabilities in (b) and (c) would be equal.

	Bachelor's	Master's	Professional	Doctorate	Total
Female	0.3883	0.1673	0.0212	0.0108	0.5876
Male	0.2751	0.1082	0.0183	0.0108	0.4124
Total	0.6633	0.2755	0.0395	0.0216	1.0000

4.123. Let M be the event "the person is a man" and B be "the person earned a bachelor's degree." **(a)** $P(M) = \frac{991}{2403} \doteq 0.4124$. Or take the answer from part (b) of the previous exercise and subtract from 1. **(b)** $P(B \mid M) = \frac{661/2403}{991/2403} = \frac{661}{991} \doteq 0.6670$. (This is the "Bachelor's" entry from the "Male" row, divided by that row's total.) **(c)** $P(M \text{ and } B) = P(M)\,P(B \mid M) \doteq (0.4124)(0.6670) \doteq 0.2751$. This agrees with the directly computed probability: $P(M \text{ and } B) = \frac{661}{2403} \doteq 0.2751$.

4.124. Each unemployment rate is computed as shown on the right. (Alternatively, subtract the number employed from the number in the labor force, then divide that difference by the number in the labor force.) Because these rates (probabilities) are different, education level and being employed are not independent.

Did not finish HS	$1 - \frac{11,552}{12,623} \doteq 0.0848$
HS/no college	$1 - \frac{36,249}{38,210} \doteq 0.0513$
Some college	$1 - \frac{32,429}{33,928} \doteq 0.0442$
College graduate	$1 - \frac{39,250}{40,414} \doteq 0.0288$

4.125. **(a)** Add up the numbers in the first and second columns. We find that there are 186,210 thousand (i.e., over 186 million) people aged 25 or older, of which 125,175 thousand are in the labor force, so $P(L) = \frac{125,175}{186,210} \doteq 0.6722$. **(b)** $P(L \mid C) = \frac{P(L \text{ and } C)}{P(C)} = \frac{40,414}{51,568} \doteq 0.7837$. **(c)** L and C are *not* independent; if they were, the two probabilities in (a) and (b) would be equal.

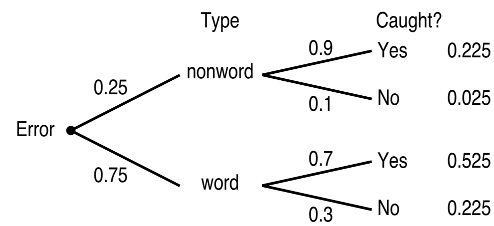
4.126. For the first probability, add up the numbers in the third column. We find that there are 119,480 thousand (i.e., over 119 million) employed people aged 25 or older. Therefore, $P(C \mid E) = \frac{P(C \text{ and } E)}{P(E)} = \frac{39{,}250}{119{,}480} \doteq 0.3285$.

For the second probability, we use the total number of college graduates in the population: $P(E \mid C) = \frac{P(C \text{ and } E)}{P(C)} = \frac{39{,}250}{51{,}568} \doteq 0.7611$.

4.127. The population includes retired people who have left the labor force. Retired persons are more likely than other adults to have not completed high school; consequently, a relatively large number of retired persons fall in the "did not finish high school" category.

Note: *Details of this lurking variable can be found in the Current Population Survey annual report on "Educational Attainment in the United States." For 2006, this report says that among the 65-and-over population, about 24.8% have not completed high school, compared to about 19.3% of the under-65 group.*

4.128. Tree diagram at right. The numbers on the right side of the tree are found by the multiplication rule; for example, P("nonword error" and "caught") $= P(N \text{ and } C) = P(N)\,P(C \mid N) = (0.25)(0.9) = 0.225$. A proofreader should catch about $0.225 + 0.525 = 0.75 = 75\%$ of all errors.



4.129. **(a)** Her brother has type aa, and he got one allele from each parent. But neither parent is albino, so neither could be type aa. **(b)** The table on the right shows the possible combinations, each of which is equally likely, so $P(aa) = 0.25$, $P(Aa) = 0.5$, and $P(AA) = 0.25$. **(c)** Beth is either AA or Aa, and $P(AA \mid \text{not } aa) = \frac{0.25}{0.75} = \frac{1}{3}$, while $P(Aa \mid \text{not } aa) = \frac{0.50}{0.75} = \frac{2}{3}$.

	A	a
A	AA	Aa
a	Aa	aa

4.130. **(a)** If Beth is Aa, then the first table on the right gives the (equally likely) allele combinations for a child, so P(child is non-albino | Beth is Aa) $= \frac{1}{2}$. If Beth is AA, then as the second table shows, their child will definitely be type Aa (and non-albino), so P(child is non-albino | Beth is AA) $= 1$. **(b)** We have:

	A	a
a	Aa	aa
a	Aa	aa

A	A
Aa	Aa
Aa	Aa

$$\begin{aligned} P(\text{child is non-albino}) &= P(\text{child } Aa \text{ and Beth } Aa) + P(\text{child } Aa \text{ and Beth } AA) \\ &= P(\text{Beth } Aa)\,P(\text{child } Aa \mid \text{Beth } Aa) + P(\text{Beth } AA)\,P(\text{child } Aa \mid \text{Beth } AA) \\ &= \tfrac{2}{3} \cdot \tfrac{1}{2} + \tfrac{1}{3} \cdot 1 = \tfrac{2}{3} \end{aligned}$$

Therefore, $P(\text{Beth is } Aa \mid \text{child is } Aa) = \frac{1/3}{2/3} = \frac{1}{2}$.

4.131. Let C be the event "Toni is a carrier," T be the event "Toni tests positive," and D be "her son has DMD." We have $P(C) = \frac{2}{3}$, $P(T \mid C) = 0.7$, and $P(T \mid C^c) = 0.1$. Therefore, $P(T) = P(T \text{ and } C) + P(T \text{ and } C^c) = P(C)\,P(T \mid C) + P(C^c)\,P(T \mid C^c) = \left(\frac{2}{3}\right)(0.7) + \left(\frac{1}{3}\right)(0.1) = 0.5$, and:

$$P(C \mid T) = \frac{P(T \text{ and } C)}{P(C)} = \frac{(2/3)(0.7)}{0.5} = \frac{14}{15} \doteq 0.9333$$

4.132. The value -1 should occur about 30% of the time (that is, the proportion should be close to 0.3).

4.133. The mean of X is $\mu_X = (-1)(0.3) + (2)(0.7) = 1.1$, so the mean if many such values will be close to 1.1.

4.134. **(a)** $\mu_X = (1)(0.4) + (2)(0.6) = 1.6$ and $\sigma_X^2 = (1-1.6)^2(0.4) + (2-1.6)^2(0.6) = 0.24$, so $\sigma_X = \sqrt{0.24} \doteq 0.4899$. **(b)** The mean is $\mu_Y = 3\mu_X - 2 = 2.8$. The variance is $\sigma_Y^2 = 9\sigma_X^2 = 2.16$, and the standard deviation is $\sigma_Y = \sqrt{2.16} \doteq 1.4697$ (this can also be found as $3\sigma_X$). **(c)** The first computation used Rule 1 for means. The second computation used Rule 1 for variances and standard deviations.

4.135. **(a)** Because the possible values of X are 1 and 2, the possible values of Y are $3 \cdot 1^2 - 2 = 1$ (with probability 0.4) and $3 \cdot 2^2 - 2 = 10$ (with probability 0.6). **(b)** $\mu_Y = (1)(0.4) + (10)(0.6) = 6.4$ and $\sigma_Y^2 = (1-6.4)^2(0.4) + (10-6.4)^2(0.6) = 19.44$, so $\sigma_X = \sqrt{19.44} \doteq 4.4091$. **(c)** Those rule are for transformations of the form $aX + b$, not $aX^2 + b$.

4.136. **(a)** A and B are disjoint. (If A happens, B did not.) **(b)** A and B are independent. (A concerns the first roll, B the second.) **(c)** A and B are independent. (A concerns the second roll, B the first.) **(d)** A and B are neither disjoint nor independent. (If A happens, then so does B.)

4.137. **(a)** $P(A) = \frac{5}{36}$ and $P(B) = \frac{10}{36} = \frac{5}{18}$. **(b)** $P(A) = \frac{5}{36}$ and $P(B) = \frac{21}{36} = \frac{7}{12}$. **(c)** $P(A) = \frac{15}{36} = \frac{5}{12}$ and $P(B) = \frac{10}{36} = \frac{5}{18}$. **(d)** $P(A) = \frac{15}{36} = \frac{5}{12}$ and $P(B) = \frac{10}{36} = \frac{5}{18}$.

4.138. **(a)** The mean is $\mu_X = (2)(0.3) + (3)(0.4) + (4)(0.3) = 3$. The variance is

$$\sigma_X^2 = (2-3)^2(0.3) + (3-3)^2(0.4) + (4-3)^2(0.3) = 0.6,$$

Value of X	2	3	4
Probability	0.3	0.4	0.3
(b) & (c)	p	$1-2p$	p

so the standard deviation is $\sigma_X = \sqrt{0.6} = 0.7746$. **(b)** & **(c)** To achieve a mean of 3 with possible values 2, 3, and 4, the distribution must be symmetric; that is, the probability at 2 must equal the probability at 4 (so that 3 would be the balance point of the distribution). Let p be the probability assigned to 2 (and also to 4) in the new distribution. A larger standard deviation is achieved when $p > 0.3$, and a smaller standard deviation arises when $p < 0.3$. In either case, the new standard deviation is $\sqrt{2p}$.

4.139. For each bet, the mean is the winning probability times the winning payout, plus the losing probability times $-\$10$. These are summarized in the table on the right; all mean payoffs equal \$0.

Point	Expected Payoff
4 or 10	$\frac{1}{3}(+\$20) + \frac{2}{3}(-\$10) = 0$
5 or 9	$\frac{2}{5}(+\$15) + \frac{3}{5}(-\$10) = 0$
6 or 8	$\frac{5}{11}(+\$12) + \frac{6}{11}(-\$10) = 0$

Note: *Alternatively, we can find the mean amount of money we have at the end of the bet. For example, if the point is 4 or 10, we end with either \$30 or \$0, and our expected ending amount is* $\frac{1}{3}(\$30) + \frac{2}{3}(\$0) = \$10$ *—equal to the amount of the bet.*

4.140. $P(\text{A}) = P(\text{B}) = \cdots = P(\text{F}) = \frac{0.72}{6} = 0.12$ and $P(1) = \cdots = P(8) = \frac{1-0.72}{8} = 0.035$.

4.141. **(a)** All probabilities are greater than or equal to 0, and their sum is 1. **(b)** Let R_1 be Taster 1's rating and R_2 be Taster 2's rating. Add the probabilities on the diagonal (upper left to lower right): $P(R_1 = R_2) = 0.03 + 0.07 + 0.25 + 0.20 + 0.06 = 0.61$. **(c)** $P(R_1 > 3) = 0.39$ (the sum of the ten numbers in the bottom two rows), and $P(R_2 > 3) = 0.39$ (the sum of the ten numbers in the right two columns).

4.142. As $\sigma_{a+bX} = b\sigma_X$ and $\sigma_{c+dY} = d\sigma_Y$, we need $b = \frac{100}{106}$ and $d = \frac{100}{109}$. With these choices for b and d, we have $\mu_{a+bX} = a + b\mu_X \doteq a + 419.8113$, so $a \doteq 80.1887$, and $\mu_{c+dX} = c + d\mu_Y \doteq c + 519.2661$, so $c \doteq -19.2661$.

4.143. This is the probability of 19 (independent) losses, followed by a win; by the multiplication rule, this is $0.994^{19} \cdot 0.006 \doteq 0.005352$.

4.144. **(a)** $P(\text{win the jackpot}) = \left(\frac{1}{20}\right)\left(\frac{8}{20}\right)\left(\frac{1}{20}\right) = 0.001$. **(b)** The other symbol can show up on the middle wheel, with probability $\left(\frac{1}{20}\right)\left(\frac{12}{20}\right)\left(\frac{1}{20}\right) = 0.0015$, or on either of the outside wheels, with probability $\left(\frac{19}{20}\right)\left(\frac{8}{20}\right)\left(\frac{1}{20}\right) = 0.019$. Therefore, combining all three cases, we have $P(\text{exactly two bells}) = 0.0015 + 2 \cdot 0.019 = 0.0395$.

4.145. With B, M, and D representing the three kinds of degrees, and W meaning the degree recipient was a woman, we have been given:

$$P(B) = 0.71, \qquad P(M) = 0.23, \qquad P(D) = 0.06,$$
$$P(W \mid B) = 0.44, \quad P(W \mid M) = 0.41, \quad P(W \mid D) = 0.30.$$

Therefore, we find

$$\begin{aligned} P(W) &= P(W \text{ and } B) + P(W \text{ and } M) + P(W \text{ and } D) \\ &= P(B)\,P(W \mid B) + P(M)\,P(W \mid M) + P(D)\,P(W \mid D) = 0.4247, \end{aligned}$$

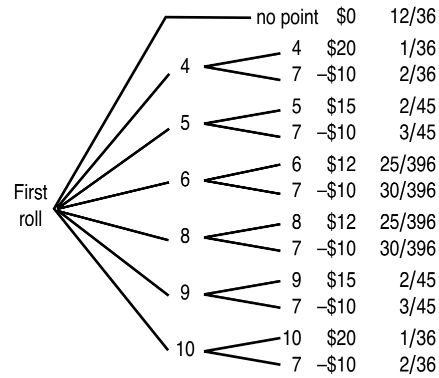
so:

$$P(B \mid W) = \frac{P(B \text{ and } W)}{P(W)} = \frac{P(B)\,P(W \mid B)}{P(W)} = \frac{0.3124}{0.4247} \doteq 0.7356$$

4.146. The table shows conditional distributions given T (public or private institution) and given Y (two-year or four-year institution). The four numbers in the "Given T" group are found by dividing each number in the table by the *column* total (so each column sums to 1). In the "Given Y" group, we divide each number by the *row* total (so each row sums to 1). For example: $P(\text{two-year} \mid \text{Public}) = \frac{639}{639 + 1061} \doteq 0.3759$, $P(\text{two-year} \mid \text{Private}) = \frac{1894}{1894 + 622} \doteq 0.7528$, and $P(\text{Public} \mid \text{two-year}) = \frac{639}{639 + 1894} \doteq 0.2523$.

	Given T		Given Y	
	Public	Private	Public	Private
Two-year	0.3759	0.7528	0.2523	0.7477
4-year	0.6241	0.2472	0.6304	0.3696

4.147. $P(\text{no point is established}) = \frac{12}{36} = \frac{1}{3}$.



In Exercise 4.139, the probabilities of winning each odds bet were given as $\frac{1}{3}$ for 4 and 10, $\frac{2}{5}$ for 5 and 9, and $\frac{5}{11}$ for 6 and 8. This tree diagram can get a bit large (and crowded). In the diagram shown on the right, the probabilities are omitted from the individual branches. The probability of winning an odds bet on 4 or 10 (with a net payout of \$20) is $\left(\frac{3}{36}\right)\left(\frac{1}{3}\right) = \frac{1}{36}$. Losing that odds bet costs \$10, and has probability $\left(\frac{3}{36}\right)\left(\frac{2}{3}\right) = \frac{2}{36}$ (or $\frac{1}{18}$). Similarly, the probability of winning an odds bet on 5 or 9 is $\left(\frac{4}{36}\right)\left(\frac{2}{5}\right) = \frac{2}{45}$, and the probability of losing that bet is $\left(\frac{4}{36}\right)\left(\frac{3}{5}\right) = \frac{3}{45}$ (or $\frac{1}{15}$). For an odds bet on 6 or 8, we win \$12 with probability $\left(\frac{5}{36}\right)\left(\frac{5}{11}\right) = \frac{25}{396}$, and lose \$10 with probability $\left(\frac{5}{36}\right)\left(\frac{6}{11}\right) = \frac{30}{396}$ (or $\frac{5}{66}$).

To confirm that this game is fair, one can multiply each payoff by its probability then add up all of those products. More directly, because each individual odds bet is fair (as was shown in the solution to Exercise 4.139), one can argue that taking the odds bet whenever it is available must be fair.

4.148. Student findings will depend on how much they explore the Web site. Individual growth charts include weight-for-age, height-for-age, weight-for-length, head circumference-for-age, and body mass index-for-age.

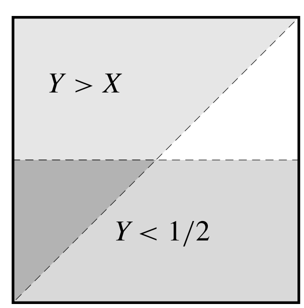
4.149. Let R_1 be Taster 1's rating and R_2 be Taster 2's rating. $P(R_1 = 3) = 0.01 + 0.05 + 0.25 + 0.05 + 0.01 = 0.37$, so:

$$P(R_2 > 3 \mid R_1 = 3) = \frac{P(R_2 > 3 \text{ and } R_1 = 3)}{P(R_1 = 3)} = \frac{0.05 + 0.01}{0.37} \doteq 0.1622$$

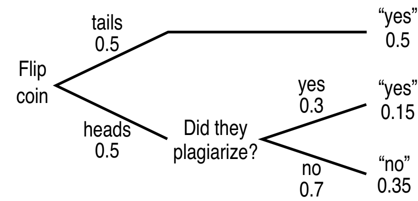
4.150. Note first that $P(A) = \frac{1}{2}$ and $P(B) = \frac{2}{4} = \frac{1}{2}$. Now $P(B \text{ and } A) = P(\text{both coins are heads}) = 0.25$, so $P(B \mid A) = \dfrac{P(B \text{ and } A)}{P(A)} = \frac{0.25}{0.5} = 0.5 = P(B)$.

4.151. The event $\{Y < 1/2\}$ is the bottom half of the square, while $\{Y > X\}$ is the upper left triangle of the square. They overlap in a triangle with area 1/8, so:

$$P(Y < \tfrac{1}{2} \mid Y > X) = \frac{P(Y < \frac{1}{2} \text{ and } Y > X)}{P(Y > X)} = \frac{1/8}{1/2} = \frac{1}{4}$$



4.152. The response will be "no" with probability $(0.5)(0.7) = 0.35$. (That is, of the 70% who have not plagiarized, half will say "no.")



If the probability of plagiarism were 0.2, then $P(\text{student answers "no"}) = (0.5)(0.8) = 0.4$. (Of the 80% who have not plagiarized, half say "no.")

If 39% of students surveyed answered "no," then we estimate that $2 \cdot 39\% = 78\%$ have *not* plagiarized, so about 22% have plagiarized.

Chapter 5 Solutions

5.1. The population is iPhone users (or iPhone users who use the AppsFire service). The statistic is an average of 65 apps per device. Likely values will vary, in part based on how many apps are on student phones (which they might consider "typical").

5.2. With $\mu = 240$, $\sigma = 18$, and $n = 36$, we have mean $\mu_{\bar{x}} = \mu = 240$ and standard deviation $\sigma_{\bar{x}} = \sigma/\sqrt{n} = 3$.

5.3. When $n = 144$, the mean is $\mu_{\bar{x}} = \mu = 240$ (unchanged), and the standard deviation is $\sigma_{\bar{x}} = \sigma/\sqrt{n} = 1.5$. Increasing n does not change $\mu_{\bar{x}}$ but decreases $\sigma_{\bar{x}}$, the variability of the sampling distribution. (In this case, because n was increased by a factor of 4, $\sigma_{\bar{x}}$ was halved.)

5.4. When $n = 144$, $\sigma_{\bar{x}} = \sigma/\sqrt{144} = 18/12 = 1.5$. The sampling distribution of $\bar{x}$ is approximately $N(240, 1.5)$, so about 95% of the time, $\bar{x}$ is between 237 and 243.

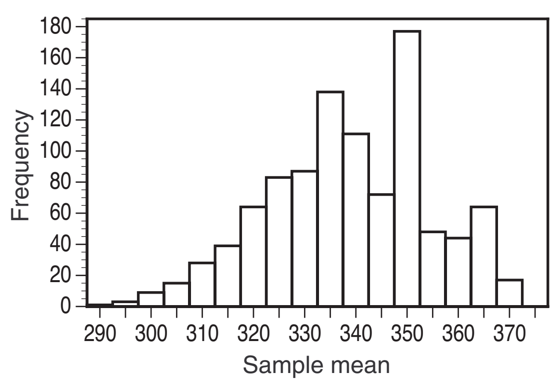
5.5. When $n = 1296$, $\sigma_{\bar{x}} = \sigma/\sqrt{1296} = 18/36 = 0.5$. The sampling distribution of $\bar{x}$ is approximately $N(240, 0.5)$, so about 95% of the time, $\bar{x}$ is between 239 and 241.

5.6. With $\sigma/\sqrt{50} \doteq 3.54$, we have $P(\bar{x} < 28) = P\left(\frac{\bar{x}-25}{3.54} < \frac{28-25}{3.54}\right) \doteq P(Z < 0.85) \doteq 0.8023$.

5.7. **(a)** Either change "variance" to "standard deviation" (twice), or change the formula at the end to $10^2/30$. **(b)** Standard deviation decreases with increasing sample size. **(c)** $\mu_{\bar{x}}$ always equals μ, regardless of the sample size.

5.8. **(a)** The distribution of $\bar{x}$ is approximately Normal. (The distribution of observed values—that is, the population distribution—is unaffected by the sample size.) **(b)** $\bar{x}$ is within $\mu \pm 2\sigma/\sqrt{n}$ about 95% of the time. **(c)** The (distribution of the) sample mean $\bar{x}$ is approximately Normal. (μ is not random; it is just a number, albeit typically an unknown one.)

5.9. **(a)** $\mu = 3388/10 = 338.8$. **(b)** The scores will vary depending on the starting row. The smallest and largest possible means are 290 and 370. **(c)** Answers will vary. Shown on the right is a histogram of the (exact) sampling distribution. With a sample size of only 3, the distribution is noticeably non-Normal. **(d)** The center of the exact sampling distribution is μ, but with only 10 values of $\bar{x}$, this might not be true for student histograms.



Note: *This histograms were found by considering all 1000 possible samples.*

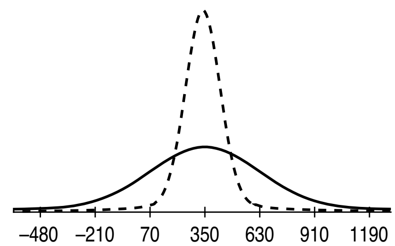
5.10. **(a)** $\sigma_{\bar{x}} = \sigma/\sqrt{200} \doteq 0.08132$. **(b)** With $n = 200$, $\bar{x}$ will be within ± 0.16 (about 10 minutes) of $\mu = 7.02$ hours. **(c)** $P(\bar{x} \le 6.9) = P\left(Z \le \frac{6.9-7.02}{0.08132}\right) \doteq P(Z \le -1.48) \doteq 0.0694$.

5.11. **(a)** With $n = 200$, the 95% probability range was about ± 10 minutes, so need a larger sample size. (Specifically, to halve the range, we need to roughly quadruple the sample size.) **(b)** We need $2\sigma_{\bar{x}} = \frac{5}{60}$, so $\sigma_{\bar{x}} \doteq 0.04167$. **(c)** With $\sigma = 1.15$, we have $\sqrt{n} = \frac{1.15}{0.04167} = 27.6$, so $n = 761.76$—use 762 students.

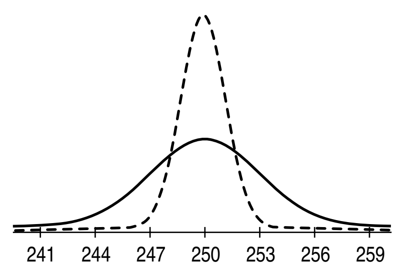
5.12. **(a)** The standard deviation is $\sigma/\sqrt{10} = 280/\sqrt{10} \doteq 88.5438$ seconds. **(b)** In order to have $\sigma/\sqrt{n} = 15$ seconds, we need $\sqrt{n} = \frac{280}{15}$, so $n \doteq 348.4$—use $n = 349$.

5.13. Mean $\mu = 250$ ml and standard deviation $\sigma/\sqrt{6} = 3/\sqrt{6} \doteq 1.2247$ ml.

5.14. **(a)** For this exercise, bear in mind that the actual distribution for a single song length is definitely *not* Normal; in particular, a Normal distribution with mean 350 seconds and standard deviation 280 seconds extends well below 0 seconds. The Normal curve for $\bar{x}$ should be taller by a factor of $\sqrt{10}$ and skinnier by a factor of $1/\sqrt{10}$ (although that technical detail will likely be lost on most students). **(b)** Using a $N(350, 280)$ distribution, $1 - P(331 < X < 369) \doteq 1 - P(-0.07 < Z < 0.07) \doteq 0.9442$. **(c)** Using a $N(350, 88.5438)$ distribution, $1 - P(331 < X < 369) \doteq 1 - P(-0.21 < Z < 0.21) \doteq 0.8336$.



5.15. In Exercise 5.13, we found that $\sigma_{\bar{x}} \doteq 1.2247$ ml, so $\bar{x}$ has a $N(250 \text{ ml}, 1.2247 \text{ ml})$ distribution. **(a)** On the right. The Normal curve for $\bar{x}$ should be taller by a factor of $\sqrt{6}$ and skinnier by a factor of $1/\sqrt{6}$. **(b)** The probability that a single can's volume differs from the target by at least 1 ml—one-third of a standard deviation—is $1 - P(-0.33 < Z < 0.33) = 0.7414$. **(c)** The probability that $\bar{x}$ is at least 1 ml from the target is



$$1 - P(249 < \bar{x} < 251) = 1 - P(-0.82 < Z < 0.82) = 0.4122.$$

5.16. For the population distribution (the number of friends of a randomly chosen individual), $\mu = 130$ and $\sigma = 85$ friends. **(a)** For the total number of friends for a sample of $n = 30$ users, the mean is $n\mu = 3900$ and the standard deviation is $\sigma\sqrt{n} \doteq 465.56$ friends. **(b)** For the mean number of friends, the mean is $\mu = 130$ and the standard deviation is $\sigma/\sqrt{n} \doteq 15.519$ friends. **(c)** $P(\bar{x} > 140) = P\left(Z > \frac{140-130}{15.519}\right) \doteq P(Z > 0.64) = 0.2611$ (software: 0.2597).

5.17. **(a)** $\bar{x}$ is not systematically higher than or lower than μ; that is, it has no particular tendency to underestimate or overestimate μ. (In other words, it is "just right" on the average.) **(b)** With large samples, $\bar{x}$ is more likely to be close to μ because with a larger sample comes more information (and therefore less uncertainty).

5.18. **(a)** $P(X \geq 23) \doteq P\left(Z \geq \frac{23-19.2}{5.1}\right) = P(Z \geq 0.75) = 0.2266$ (with software: 0.2281). Because ACT scores are reported as whole numbers, we might instead compute $P(X \geq 22.5) \doteq P(Z \geq 0.65) = 0.2578$ (software: 0.2588). **(b)** $\mu_{\bar{x}} = 19.2$ and $\sigma_{\bar{x}} = \sigma/\sqrt{25} = 1.02$. **(c)** $P(\bar{x} \geq 23) \doteq P\left(Z \geq \frac{23-19.2}{1.02}\right) = P(Z \geq 3.73) = 0.0001$. (In this case, it is not appropriate to find $P(\bar{x} \geq 22.5)$, unless $\bar{x}$ is rounded to the nearest whole number.) **(d)** Because individual scores are only roughly Normal, the answer to (a) is approximate. The answer to (c) is also approximate but should be more accurate because $\bar{x}$ should have a distribution that is closer to Normal.

5.19. **(a)** $\mu_{\bar{x}} = 0.5$ and $\sigma_{\bar{x}} = \sigma/\sqrt{50} = 0.7/\sqrt{50} \doteq 0.09899$. **(b)** Because this distribution is only approximately Normal, it would be quite reasonable to use the 68–95–99.7 rule to give a rough estimate: 0.6 is about one standard deviation above the mean, so the probability should be about 0.16 (half of the 32% that falls outside ± 1 standard deviation). Alternatively, $P(\bar{x} > 0.6) \doteq P\left(Z > \frac{0.6-0.5}{0.09899}\right) = P(Z > 1.01) = 0.1562$.

5.20. **(a)** $\mu = (4)(0.33) + (3)(0.24) + (2)(0.18) + (1)(0.16) + (0)(0.09) = 2.56$ and

$$\sigma^2 = (4-2.56)^2(0.33) + (3-2.56)^2(0.24) + (2-2.56)^2(0.18) + (1-2.56)^2(0.16) + (0-2.56)^2(0.09) = 1.7664,$$

so $\sigma = \sqrt{1.7664} \doteq 1.3291$. **(b)** $\mu_{\bar{x}} = \mu = 2.56$ and $\sigma_{\bar{x}} = \sigma/\sqrt{50} \doteq 0.1880$. **(c)** $P(X \geq 3) = 0.33 + 0.24 = 0.57$, and $P(\bar{x} \geq 3) \doteq P\left(Z \geq \frac{3-2.56}{0.1880}\right) = P(Z \geq 2.34) = 0.0096$.

5.21. Let X be Sheila's measured glucose level. **(a)** $P(X > 140) = P(Z > 1.5) = 0.0668$. **(b)** If $\bar{x}$ is the mean of three measurements (assumed to be independent), then $\bar{x}$ has a $N(125, 10/\sqrt{3})$ or $N(125 \text{ mg/dl}, 5.7735 \text{ mg/dl})$ distribution, and $P(\bar{x} > 140) = P(Z > 2.60) = 0.0047$.

5.22. **(a)** $\mu_X = (\$500)(0.001) = \$0.50$ and $\sigma_X = \sqrt{249.75} \doteq \$15.8035$. **(b)** In the long run, Joe makes about 50 cents for each \$1 ticket. **(c)** If $\bar{x}$ is Joe's average payoff over a year, then $\mu_{\bar{x}} = \mu = \$0.50$ and $\sigma_{\bar{x}} = \sigma_X/\sqrt{104} \doteq \$1.5497$. The central limit theorem says that $\bar{x}$ is approximately Normally distributed (with this mean and standard deviation). **(d)** Using this Normal approximation, $P(\bar{x} > \$1) \doteq P(Z > 0.32) = 0.3745$ (software: 0.3735).

Note: *Joe comes out ahead if he wins at least once during the year. This probability is easily computed as* $1 - (0.999)^{104} \doteq 0.0988$. *The distribution of* $\bar{x}$ *is different enough from a Normal distribution so that answers given by the approximation are not as accurate in this case as they are in many others.*

5.23. The mean of three measurements has a $N(125 \text{ mg/dl}, 5.7735 \text{ mg/dl})$ distribution, and $P(Z > 1.645) = 0.05$ if Z is $N(0, 1)$, so $L = 125 + 1.645 \cdot 5.7735 \doteq 134.5$ mg/dl.

5.24. $\bar{x}$ is approximately Normal with mean 1.3 and standard deviation $1.5/\sqrt{200} \doteq 0.1061$ flaws/yd^2, so $P(\bar{x} > 2) \doteq P(Z > 6.6) = 0$ (essentially).

5.25. If W is total weight, and $\bar{x} = W/25$, then:

$$P(W > 5200) = P(\bar{x} > 208) \doteq P\left(Z > \tfrac{208-190}{5/\sqrt{25}}\right) = P(Z > 2.57) = 0.0051$$

5.26. **(a)** Although the probability of having to pay for a total loss for one or more of the 12 policies is very small, if this were to happen, it would be financially disastrous. On the other hand, for thousands of policies, the law of large numbers says that the average claim on many policies will be close to the mean, so the insurance company can be assured that the premiums they collect will (almost certainly) cover the claims. **(b)** The central limit theorem says that, in spite of the skewness of the population distribution, the average loss among 10,000 policies will be approximately Normally distributed with mean \$250 and standard deviation $\sigma/\sqrt{10{,}000} = \$1000/100 = \$10$. Since \$275 is 2.5 standard deviations above the mean, the probability of seeing an average loss over \$275 is about 0.0062.

5.27. **(a)** The mean of six untreated specimens has a standard deviation of $2.2/\sqrt{6} \doteq 0.8981$ lbs, so $P(\bar{x}_u > 50) = P\left(Z > \frac{50-57}{0.8981}\right) = P(Z > -7.79)$, which is basically 1. **(b)** $\bar{x}_u - \bar{x}_t$ has mean $57 - 30 = 27$ lbs and standard deviation $\sqrt{2.2^2/6 + 1.6^2/6} \doteq 1.1106$ lbs, so $P(\bar{x}_u - \bar{x}_t > 25) = P\left(Z > \frac{25-27}{1.1106}\right) = P(Z > -1.80) \doteq 0.9641$.

5.28. **(a)** The central limit theorem says that the sample means will be roughly Normal. Note that the distribution of individual scores cannot have extreme outliers because all scores are between 1 and 7. **(b)** For *Journal* scores, $\bar{y}$ has mean 4.8 and standard deviation $1.5/\sqrt{28} \doteq 0.2835$. For *Enquirer* scores, $\bar{x}$ has mean 2.4 and standard deviation $1.6/\sqrt{28} \doteq 0.3024$. **(c)** $\bar{y} - \bar{x}$ has (approximately) a Normal distribution with mean 2.4 and standard deviation $\sqrt{1.5^2/28 + 1.6^2/28} \doteq 0.4145$. **(d)** $P(\bar{y} - \bar{x} \ge 1) = P\left(Z \ge \frac{1-2.4}{0.4145}\right) = P(Z \ge -3.38) \doteq 0.9996$.

5.29. **(a)** $\bar{y}$ has a $N(\mu_Y, \sigma_Y/\sqrt{m})$ distribution and $\bar{x}$ has a $N(\mu_X, \sigma_X/\sqrt{n})$ distribution. **(b)** $\bar{y} - \bar{x}$ has a Normal distribution with mean $\mu_Y - \mu_X$ and standard deviation $\sqrt{\sigma_Y^2/m + \sigma_X^2/n}$.

5.30. We have been given $\mu_X = 9\%$, $\sigma_X = 19\%$, $\mu_Y = 11\%$, $\sigma_Y = 17\%$, and $\rho = 0.6$. **(a)** Linda's return $R = 0.7X + 0.3Y$ has mean $\mu_R = 0.7\mu_X + 0.3\mu_Y = 9.6\%$ and standard deviation $\sigma_R = \sqrt{(0.7\sigma_X)^2 + (0.3\sigma_Y)^2 + 2\rho(0.7\sigma_X)(0.3\sigma_Y)} \doteq 16.8611\%$. **(b)** $\overline{R}$, the average return over 20 years, has approximately a Normal distribution with mean 9.6% and standard deviation $\sigma_R/\sqrt{20} \doteq 3.7703\%$, so $P(\overline{R} < 5\%) \doteq P(Z < -1.22) \doteq 0.1112$. **(c)** After a 12% gain in the first year, Linda would have \$1120; with a 6% gain in the second year, her portfolio would be worth \$1187.20. By contrast, two years with a 9% return would make her portfolio worth \$1188.10.

Note: *As the text suggests, the appropriate average for this situation is (a variation on) the* geometric mean, *computed as* $\sqrt{(1.12)(1.06)} - 1 \doteq 8.9587\%$. *Generally, if the sequence of annual returns is* $r_1, r_2, \ldots, r_k$ *(expressed as decimals), the mean return is*

$\sqrt[k]{(1+r_1)(1+r_2)\cdots(1+r_k)}-1$. *It can be shown that the geometric mean is always smaller than the arithmetic mean, unless all the returns are the same.*

5.31. The total height H of the four rows has a Normal distribution with mean $4 \times 8 = 32$ inches and standard deviation $0.1\sqrt{4} = 0.2$ inch. $P(H < 31.5 \text{ or } H > 32.5) = 1 - P(31.5 < H < 32.5) = 1 - P(-2.50 < Z < 2.50) = 1 - 0.9876 = 0.0124$.

5.32. $n = 250$ (the sample size), $\hat{p} = 45\% = 0.45$, and $X = n\hat{p} = 112.5$. (Because X must be a whole number, it was either 112 or 113, and the reported value of $\hat{p}$ was rounded.)

5.33. (a) $n = 1500$ (the sample size). **(b)** The "Yes" count seems like the most reasonable choice, but either count is defensible. **(c)** $X = 825$ (or $X = 675$). **(d)** $\hat{p} = \frac{825}{1500} = 0.55$ (or $\hat{p} = \frac{675}{1500} = 0.45$).

5.34. Assuming no multiple births (twins, triplets, quadruplets), we have four independent trials, each with probability of success (type O blood) equal to 0.25, so the number of children with type O blood has the $B(4, 0.25)$ distribution.

5.35. We have 15 independent trials, each with probability of success (heads) equal to 0.5, so X has the $B(15, 0.5)$ distribution.

5.36. Assuming each free-throw attempt is an independent trial, X has the $B(10, 0.8)$ distribution, and $P(X \le 4) = 0.0064$.

5.37. (a) For the $B(5, 0.4)$ distribution, $P(X = 0) = 0.0778$ and $P(X \ge 3) = 0.3174$. **(b)** For the $B(5, 0.6)$ distribution, $P(X = 5) = 0.0778$ and $P(X \le 2) = 0.3174$. **(c)** The number of "failures" in the $B(5, 0.4)$ distribution has the $B(5, 0.6)$ distribution. With 5 trials, 0 successes is equivalent to 5 failures, and 3 or more successes is equivalent to 2 or fewer failures.

5.38. (a) For the $B(100, 0.5)$ distribution, $\mu_{\hat{p}} = p = 0.5$ and $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \frac{1}{20} = 0.05$. **(b)** No; the mean and standard deviation of the sample count are both 100 times bigger. (That is, $\hat{p} = X/100$, so $\mu_{\hat{p}} = \mu_X/100$ and $\sigma_{\hat{p}} = \sigma_X/100$.)

5.39. (a) $\hat{p}$ has approximately a Normal distribution with mean 0.5 and standard deviation 0.05, so $P(0.3 < \hat{p} < 0.7) = P(-4 < Z < 4) \doteq 1$. **(b)** $P(0.35 < \hat{p} < 0.65) = P(-3 < Z < 3) \doteq 0.9974$.

Note: *For the second, the 68–95–99.7 rule would give 0.997—an acceptable answer, especially since this is an approximation anyway. For comparison, the exact answers (to four decimal places) are* $P(0.3 < \hat{p} < 0.7) \doteq 0.9999$ *or* $P(0.3 \le \hat{p} \le 0.7) \doteq 1.0000$, *and* $P(0.35 < \hat{p} < 0.65) \doteq 0.9965$ *or* $P(0.35 \le \hat{p} \le 0.65) \doteq 0.9982$. *(Notice that the "correct" answer depends on our understanding of "between.")*

5.40. (a) $P(X \ge 3) = \binom{4}{3}0.53^3 0.47 + \binom{4}{4}0.53^4 \doteq 0.3588$. **(b)** If the coin were fair, $P(X \ge 3) = \binom{4}{3}0.5^3 0.5 + \binom{4}{4}0.5^4 = 0.3125$.

5.41. **(a)** Separate flips are independent (coins have no "memory," so they do not try to compensate for a lack of tails). **(b)** Separate flips are independent (coins have no "memory," so they do not get on a "streak" of heads). **(c)** $\hat{p}$ can vary from one set of observed data to another; it is not a parameter.

5.42. **(a)** X is a *count;* $\hat{p}$ is a proportion. **(b)** The given formula is the *standard deviation* for a binomial *proportion*. The variance for a binomial count is $np(1-p)$. **(c)** The rule of thumb in the text is that np and $n(1-p)$ should both be at least 10. If p is close to 0 (or close to 1), $n = 1000$ might not satisfy this rule of thumb. (See also the solution to Exercise 5.22.)

5.43. **(a)** A $B(200, p)$ distribution seems reasonable for this setting (even though we do not know what p is). **(b)** This setting is not binomial; there is no fixed value of n. **(c)** A $B(500, 1/12)$ distribution seems appropriate for this setting. **(d)** This is not binomial, because separate cards are not independent.

5.44. **(a)** This is not binomial; X is not a count of successes. **(b)** A $B(20, p)$ distribution seems reasonable, where p (unknown) is the probability of a defective pair. **(c)** This should be (at least approximately) the $B(n, p)$ distribution, where n is the number of students in our sample, and p is the probability that a randomly-chosen student eats at least five servings of fruits and vegetables.

5.45. **(a)** C, the number caught, is $B(10, 0.7)$. M, the number missed, is $B(10, 0.3)$. **(b)** Referring to Table C, we find $P(M \geq 4) = 0.2001 + 0.1029 + 0.0368 + 0.0090 + 0.0014 + 0.0001 = 0.3503$ (software: 0.3504).

5.46. **(a)** The $B(20, 0.3)$ distribution (at least approximately). **(b)** $P(X \geq 8) = 0.2277$.

5.47. **(a)** The mean of C is $(10)(0.7) = 7$ errors caught; for M the mean is $(10)(0.3) = 3$ errors missed. **(b)** The standard deviation of C (or M) is $\sigma = \sqrt{(10)(0.7)(0.3)} \doteq 1.4491$ errors. **(c)** With $p = 0.9$, $\sigma = \sqrt{(10)(0.9)(0.1)} \doteq 0.9487$ errors; with $p = 0.99$, $\sigma \doteq 0.3146$ errors. σ decreases toward 0 as p approaches 1.

5.48. X, the number who listen to streamed music daily, has the $B(20, 0.25)$ distribution. **(a)** $\mu_X = np = 5$, and $\mu_{\hat{p}} = 0.25$. **(b)** With $n = 200$, $\mu_X = 50$ and $\mu_{\hat{p}} = 0.25$. With $n = 2000$, $\mu_X = 500$ and $\mu_{\hat{p}} = 0.25$. μ_X increases with n, while $\mu_{\hat{p}}$ does not depend on n.

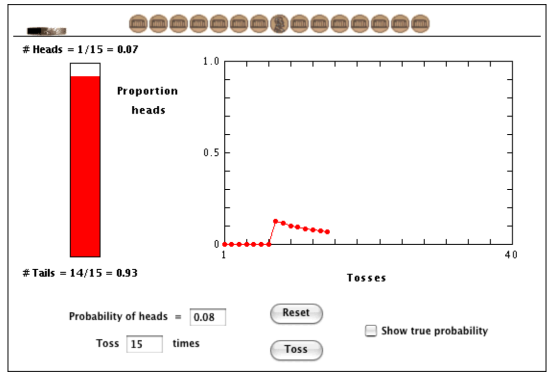
5.49. $m = 6$: $P(X \geq 6) = 0.0473$ and $P(X \geq 5) = 0.1503$.

5.50. **(a)** The population (the 75 members of the fraternity) is only 2.5 times the size of the sample. Our rule of thumb says that this ratio should be at least 20. **(b)** Our rule of thumb for the Normal approximation calls for np and $n(1-p)$ to be at least 10; we have $np = (1000)(0.002) = 2$.

5.51. The count of 5s among n random digits has a binomial distribution with $p = 0.1$. **(a)** $P(\text{at least one } 5) = 1 - P(\text{no } 5) = 1 - (0.9)^5 \doteq 0.4095$. (Or take 0.5905 from Table C and subtract from 1.) **(b)** $\mu = (40)(0.1) = 4$.

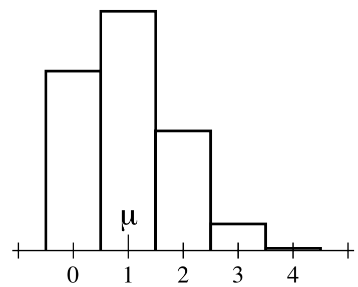
5.52. One sample of 15 flips is shown on the right. Results will vary quite a bit; Table C shows that 99.5% of the time, there will be 4 or fewer bad records in a sample of 15.

Out of 25 samples, most students should see 2 to 12 samples with no bad records. That is, N, the number of samples with no bad records, has the $B(25, 0.2683)$ distribution, and $P(2 \le N \le 12) = 0.9894$.



5.53. **(a)** $n = 4$ and $p = 1/4 = 0.25$. **(b)** The distribution is below; the histogram is on the right. **(c)** $\mu = np = 1$.

x	0	1	2	3	4
$P(X = x)$	.3164	.4219	.2109	.0469	.0039



5.54. For $\hat{p}$, $\mu = 0.52$ and $\sigma = \sqrt{p(1-p)/n} \doteq 0.01574$. As $\hat{p}$ is approximately Normally distributed with this mean and standard deviation, we find:

$$P(0.49 < \hat{p} < 0.55) \doteq P(-1.91 < Z < 1.91) \doteq 0.9438$$

(Software computation of the Normal probability gives 0.9433. Using a binomial distribution, we can also find $P(493 \le X \le 554) \doteq 0.9495$.)

5.55. Recall that $\hat{p}$ is approximately Normally distributed with mean $\mu = p$ and standard deviation $\sqrt{p(1-p)/n}$. **(a)** With $p = 0.24$, $\sigma \doteq 0.01333$, so $P(0.22 < \hat{p} < 0.26) = P(-1.50 < Z < 1.50) \doteq 0.8664$. (Software computation of the Normal probability gives 0.8666. Using a binomial distribution, we can also find $P(226 \le X \le 267) \doteq 0.8752$.) **(b)** With $p = 0.04$, $\sigma \doteq 0.00611$, so $P(0.02 < \hat{p} < 0.06) = P(-3.27 < Z < 3.27) = 0.9990$. (Using a binomial distribution, we can also find $P(21 \le X \le 62) \doteq 0.9992$.) **(c)** $P(-0.02 < \hat{p} - p < 0.02)$ increases to 1 as p gets closer to 0. (This is because σ also gets close to 0, so that $0.02/\sigma$ grows.)

5.56. When $n = 300$, the distribution of $\hat{p}$ is approximately Normal with mean 0.52 and standard deviation 0.02884 (nearly twice that in Exercise 5.54). When $n = 5000$, the standard deviation drops to 0.00707 (less than half as big as in Exercise 5.54). Therefore:

$$n = 300: \quad P(0.49 < \hat{p} < 0.55) \doteq P(-1.04 < Z < 1.04) \doteq 0.7016$$
$$n = 5000: \quad P(0.49 < \hat{p} < 0.55) \doteq P(-4.25 < Z < 4.25) \doteq 1$$

Larger samples give a better probability that $\hat{p}$ will be close to the true proportion p. (Software computation of the first Normal probability gives 0.7017; using a binomial distribution, we can also find $P(147 \le X \le 165) \doteq 0.7278$. These more accurate answers do not change our conclusion.)

5.57. **(a)** The mean is $\mu = p = 0.69$, and the standard deviation is $\sigma = \sqrt{p(1-p)/n} \doteq 0.0008444$. **(b)** $\mu \pm 2\sigma$ gives the range 68.83% to 69.17%. **(c)** This range is considerably narrower than the historical range. In fact, 67% and 70% correspond to $z = -23.7$ and $z = 11.8$—suggesting that the observed percents do not come from a $N(0.69, 0.0008444)$ distribution; that is, the population proportion has changed over time.

5.58. **(a)** $\hat{p} = \frac{294}{400} = 0.735$. **(b)** With $p = 0.8$, $\sigma_{\hat{p}} = \sqrt{(0.8)(0.2)/400} = 0.02$. **(c)** Still assuming that $p = 0.8$, we would expect that about 95% of the time, $\hat{p}$ should fall between 0.76 and 0.84. **(d)** It appears that these students prefer this type of course less than the national average. (The observed value of $\hat{p}$ is quite a bit lower than we would expect from a $N(0.8, 0.2)$ distribution, which suggests that it came from a distribution with a lower mean.)

5.59. **(a)** $\hat{p} = \frac{140}{200} = 0.7$. **(b)** We want $P(X \geq 140)$ or $P(\hat{p} \geq 0.7)$. The first can be found exactly (using a binomial distribution), or we can compute either using a Normal approximation (with or without the continuity correction). All possible answers are shown on the right. **(c)** The sample results are higher than the national percentage, but the sample was so small that such a difference could arise by chance even if the true campus proportion is the same.

			Continuity correction	
Exact prob.	Table Normal	Software Normal	Table Normal	Software Normal
0.2049	0.1841	0.1835	0.2033	0.2041

5.60. As $\sigma_{\hat{p}} = \sqrt{p(1-p)/n}$, we have $0.004^2 = (0.52)(0.48)/n$, so $n = 15{,}600$.

5.61. **(a)** $p = 1/4 = 0.25$. **(b)** $P(X \geq 10) = 0.0139$. **(c)** $\mu = np = 5$ and $\sigma = \sqrt{np(1-p)} = \sqrt{3.75} \doteq 1.9365$ successes. **(d)** No: The trials would not be independent because the subject may alter his/her guessing strategy based on this information.

5.62. **(a)** $\mu = (1200)(0.75) = 900$ and $\sigma = \sqrt{225} = 15$ students. **(b)** $P(X \geq 800) \doteq P(Z \geq -6.67) = 1$ (essentially). **(c)** $P(X \geq 951) \doteq P(Z \geq 3.4) = 0.0003$. **(d)** With $n = 1300$, $P(X \geq 951) \doteq P(Z \geq -1.54) = 0.9382$. Other answers are shown in the table on the right.

		Continuity correction	
Table Normal	Software Normal	Table Normal	Software Normal
0.9382	0.9379	0.9418	0.9417

5.63. **(a)** X, the count of successes, has the $B(900, 1/5)$ distribution, with mean $\mu_X = np = (900)(1/5) = 180$ and $\sigma_X = \sqrt{(900)(0.2)(0.8)} = 12$ successes. **(b)** For $\hat{p}$, the mean is $\mu_{\hat{p}} = p = 0.2$ and $\sigma_{\hat{p}} = \sqrt{(0.2)(0.8)/900} \doteq 0.01333$. **(c)** $P(\hat{p} > 0.24) \doteq P(Z > 3) = 0.0013$. **(d)** From a standard Normal distribution, $P(Z > 2.326) = 0.01$, so the subject must score 2.326 standard deviations above the mean: $\mu_{\hat{p}} + 2.326\sigma_{\hat{p}} = 0.2310$. This corresponds to 208 or more successes.

5.64. **(a)** M has the $B(30, 0.65)$ distribution, so $P(M = 20) = \binom{30}{20}(0.65)^{20}(0.35)^{10} \doteq 0.1502$. **(b)** $P(\text{1st woman is the 4th call}) = (0.65)^3(0.35) = 0.0961$.

5.65. **(a)** $p = \frac{23{,}772{,}494}{209{,}128{,}094} \doteq 0.1137$. **(b)** If B is the number of blacks, then B has (approximately) the $B(1200, 0.1137)$ distribution, so the mean is $np \doteq 136.4$ blacks. **(c)** $P(B \leq 100) \doteq P(Z < -3.31) = 0.0005$.

Note: *In (b), the population is at least 20 times as large as the sample, so our "rule of thumb" for using a binomial distribution is satisfied. In fact, the mean would be the same even if we could not use a binomial distribution, but we need to have a binomial distribution for part (c), so that we can approximate it with a Normal distribution—which we can safely do, because both np and n(1 − p) are much greater than 10.*

5.66. **(a)** $\binom{n}{n} = \frac{n!}{n!\,0!} = 1$. The only way to distribute n successes among n observations is for all observations to be successes. **(b)** $\binom{n}{n-1} = \frac{n!}{(n-1)!\,1!} = \frac{n \cdot (n-1)!}{(n-1)!} = n$. To distribute $n - 1$ successes among n observations, the one failure must be either observation 1, 2, 3, $\ldots, n-1$, or n. **(c)** $\binom{n}{k} = \frac{n!}{k!\,(n-k)!} = \frac{n!}{(n-k)!\,[n-(n-k)]!} = \binom{n}{n-k}$. Distributing k successes is equivalent to distributing $n - k$ failures.

5.67. Jodi's number of correct answers will have the $B(n, 0.88)$ distribution. **(a)** $P(\hat{p} \leq 0.85) = P(X \leq 85)$ is on line 1. **(b)** $P(\hat{p} \leq 0.85) = P(X \leq 212)$ is on line 2. **(c)** For a test with 400 questions, the standard deviation of $\hat{p}$ would be half as big as the standard deviation of $\hat{p}$ for a test with 100 questions: With $n = 100$, $\sigma = \sqrt{(0.88)(0.12)/100} \doteq 0.03250$; and with $n = 400$, $\sigma = \sqrt{(0.88)(0.12)/400} \doteq 0.01625$. **(d)** Yes: Regardless of p, n must be quadrupled to cut the standard deviation in half.

			Continuity correction	
Exact prob.	Table Normal	Software Normal	Table Normal	Software Normal
0.2160	0.1788	0.1780	0.2206	0.2209
0.0755	0.0594	0.0597	0.0721	0.0722

5.68. **(a)** $P(\text{first } \boxdot \text{ appears on toss } 2) = \left(\frac{5}{6}\right)\left(\frac{1}{6}\right) = \frac{5}{36}$.

(b) $P(\text{first } \boxdot \text{ appears on toss } 3) = \left(\frac{5}{6}\right)\left(\frac{5}{6}\right)\left(\frac{1}{6}\right) = \frac{25}{216}$.

(c) $P(\text{first } \boxdot \text{ appears on toss } 4) = \left(\frac{5}{6}\right)^3\left(\frac{1}{6}\right)$.

$P(\text{first } \boxdot \text{ appears on toss } 5) = \left(\frac{5}{6}\right)^4\left(\frac{1}{6}\right)$.

5.69. Y has possible values $1, 2, 3, \ldots$. $P(\text{first } \boxdot \text{ appears on toss } k) = \left(\frac{5}{6}\right)^{k-1}\left(\frac{1}{6}\right)$.

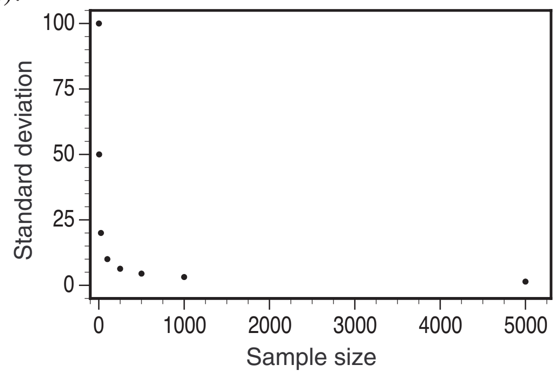
5.70. With $\mu = \$430$, $\sigma = \$140$, and $n = 500$, the distribution of $\bar{x}$ is approximately Normal with mean \$430 and $\sigma_{\bar{x}} = 140/\sqrt{500} \doteq \$6.2610$, so $P(\bar{x} > 440) \doteq P(Z > 1.60) = 0.0548$ (software: 0.0551).

5.71. **(a)** With $\sigma_{\bar{x}} = 0.08/\sqrt{3} \doteq 0.04619$, $\bar{x}$ has (approximately) a $N(123 \text{ mg}, 0.04619 \text{ mg})$ distribution. **(b)** $P(\bar{x} \geq 124) = P(Z \geq 21.65)$, which is essentially 0.

5.72. **(a)** The table of standard deviations is given below. **(b)** The graph is below on the right; it is shown as a scatterplot, but in this situation it would be reasonable to "connect the dots" because the relationship between standard deviation and sample size holds for all n. **(c)** As

n increases, the standard deviation decreases—at first quite rapidly, then more slowly (a demonstration of the law of diminishing returns).

n	$\sigma/\sqrt{n}$
1	100
4	50
25	20
100	10
250	6.32
500	4.47
1000	3.16
5000	1.41



5.73. **(a)** Out of 12 independent vehicles, the number X with one person has the $B(12, 0.755)$ distribution, so $P(X \geq 7) = 0.9503$ (using software or a calculator). **(b)** Y (the number of one-person cars in a sample of 80) has the $B(80, 0.755)$ distribution. Regardless of the approach used—Normal approximation, or exact computation using software or a calculator—$P(Y \geq 41) \doteq 1$.

5.74. This would not be surprising: Assuming that all the authors are independent (for example, none were written by siblings or married couples), we can view the 12 names as being a random sample so that the number N of occurrences of the ten most common names would have a binomial distribution with $n = 12$ and $p = 0.056$. Then $P(N = 0) = (1 - 0.056)^{12} \doteq 0.5008$.

5.75. The probability that the first digit is 1, 2, or 3 is $0.301 + 0.176 + 0.125 = 0.602$, so the number of invoices for amounts beginning with these digits should have a binomial distribution with $n = 1000$ and $p = 0.602$. More usefully, the proportion $\hat{p}$ of such invoices should have approximately a Normal distribution with mean $p = 0.602$ and standard deviation $\sqrt{p(1-p)/1000} \doteq 0.01548$, so $P(\hat{p} \leq \frac{560}{1000}) \doteq P(Z \leq -2.71) = 0.0034$. Alternate answers shown on the right.

		Continuity correction	
Table Normal	Software Normal	Table Normal	Software Normal
0.0034	0.0033	0.0037	0.0037

5.76. **(a)** If R is the number of red-blossomed plants out of a sample of 12, then $P(R = 9) = 0.2581$, using a binomial distribution with $n = 12$ and $p = 0.75$. (For Table C, use $p = 0.25$ and find $P(X = 3)$, where $X = 12 - R$ is the number of flowers with nonred blossoms.) **(b)** With $n = 120$, the mean number of red-blossomed plants is $np = 90$. **(c)** If R_2 is the number of red-blossomed plants out of a sample of 120, then $P(R_2 \geq 80) \doteq P(Z \geq -2.11) = 0.9826$. (Other possible answers are given in the table on the right.)

			Continuity correction	
Exact prob.	Table Normal	Software Normal	Table Normal	Software Normal
0.9845	0.9826	0.9825	0.9864	0.9866

5.77. If $\bar{x}$ is the average weight of 12 eggs, then $\bar{x}$ has a $N(65\text{ g}, 5/\sqrt{12}\text{ g}) = N(65\text{ g}, 1.4434\text{ g})$ distribution, and $P(\frac{755}{12} < \bar{x} < \frac{830}{12}) \doteq P(-1.44 < Z < 2.89) = 0.9231$ (software: 0.9236).

5.78. **(a)** The machine that makes the caps and the machine that applies the torque are not the same. **(b)** T (torque) is $N(7.0, 0.9)$ and S (cap strength) is $N(10.1, 1.2)$, so $T - S$ is $N(7 - 10.1, \sqrt{0.9^2 + 1.2^2}) = N(-3.1\text{ inch}\cdot\text{lb}, 1.5\text{ inch}\cdot\text{lb})$. The probability that the cap breaks is $P(T > S) = P(T - S > 0) = P(Z > 2.07) = 0.0192$ (software: 0.0194).

5.79. The center line is $\mu_{\bar{x}} = \mu = 4.25$ and the control limits are $\mu \pm 3\sigma/\sqrt{5} = 4.0689$ to 4.4311.

5.80. **(a)** $\bar{x}$ has a $N(32, 6/\sqrt{25}) = N(32, 1.2)$ distribution, and $\bar{y}$ has a $N(29, 5/\sqrt{25}) = N(29, 1)$ distribution. **(b)** $\bar{y} - \bar{x}$ has a $N(29 - 32, \sqrt{5^2/25 + 6^2/25}) \doteq N(-3, 1.5620)$ distribution. **(c)** $P(\bar{y} \geq \bar{x}) = P(\bar{y} - \bar{x} \geq 0) = P(Z \geq 1.92) = 0.0274$.

5.81. **(a)** $\hat{p}_F$ is approximately $N(0.82, 0.01921)$ and $\hat{p}_M$ is approximately $N(0.88, 0.01625)$. **(b)** When we subtract two independent Normal random variables, the difference is Normal. The new mean is the difference of the two means $(0.88 - 0.82 = 0.06)$, and the new variance is the sum of the variances $(0.01921^2 + 0.01625^2 = 0.000633)$, so $\hat{p}_M - \hat{p}_F$ is approximately $N(0.06, 0.02516)$. **(c)** $P(\hat{p}_F > \hat{p}_M) = P(\hat{p}_M - \hat{p}_F < 0) \doteq P(Z < -2.38) = 0.0087$ (software: 0.0085).

5.82. **(a)** Yes; this rule works for any random variables X and Y. **(b)** No; this rule requires that X and Y be independent. The incomes of two married people are certainly not independent, as they are likely to be similar in many characteristics that affect income (for example, educational background).

5.83. For each step of the random walk, the mean is $\mu = (1)(0.6) + (-1)(0.4) = 0.2$, the variance is $\sigma^2 = (1 - 0.2)^2(0.6) + (-1 - 0.2)^2(0.4) = 0.96$, and the standard deviation is $\sigma = \sqrt{0.96} \doteq 0.9798$. Therefore, $Y/500$ has approximately a $N(0.2, 0.04382)$ distribution, and $P(Y \geq 200) = P(\frac{Y}{500} \geq 0.4) \doteq P(Z \geq 4.56) \doteq 0$.

Note: *The number R of right-steps has a binomial distribution with $n = 500$ and $p = 0.6$. $Y \geq 200$ is equivalent to taking at least 350 right-steps, so we can also compute this probability as $P(R \geq 350)$, for which software gives the exact value 0.00000215....*

Chapter 6 Solutions

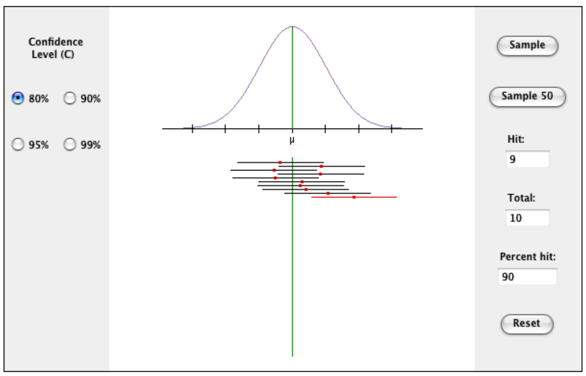
6.1. $\sigma_{\bar{x}} = \sigma/\sqrt{25} = \frac{\$2.50}{5} = \$0.50$.

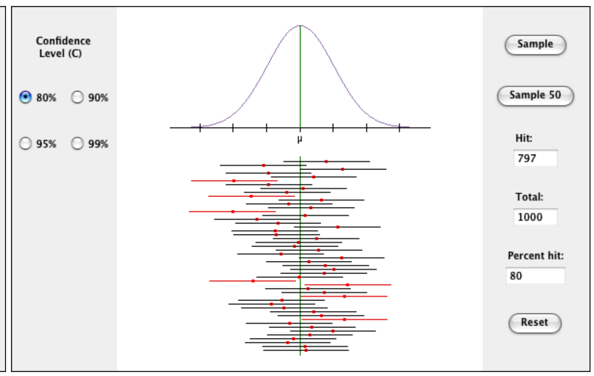
6.2. Take two standard deviations: $2 \cdot \frac{\$2.50}{5} = 2 \cdot \$0.50 = \$1.00$.

6.3. As in the previous solution, take two standard deviations: \$1.00.

Note: *This is the whole idea behind a confidence interval: Probability tells us that $\bar{x}$ is usually close to μ. That is equivalent to saying that μ is usually close to $\bar{x}$.*

6.4. Shown below are sample output screens for (a) 10 and (b) 1000 SRSs. In 99.4% of all repetitions of part (a), students should see between 5 and 10 hits (that is, at least 5 of the 10 SRSs capture the true mean μ). Out of 1000 80% confidence intervals, nearly all students will observe between 76% and 84% capturing the mean.





6.5. The standard error is $s_{\bar{x}} = \frac{\sigma}{\sqrt{100}} = 0.3$, and the 95% confidence interval for μ is

$$87.3 \pm 1.96\left(\frac{3}{\sqrt{100}}\right) = 87.3 \pm 0.588 = 86.712 \text{ to } 87.888$$

6.6. A 99% confidence interval would have a larger margin of error; a wider interval is needed in order to be more confident that the interval includes the true mean. The 99% confidence interval for μ is

$$87.3 \pm 2.576\left(\frac{3}{\sqrt{100}}\right) = 87.3 \pm 0.7728 = 86.527 \text{ to } 88.073$$

6.7. $n = \left(\frac{(1.96)(12{,}000)}{1000}\right)^2 \doteq 553.19$—take $n = 554$.

6.8. In the previous exercise, we found that $n = 554$ would give a margin of error of \$1000. The margin of error ($1.96 \times \frac{\$12{,}000}{\sqrt{n}}$) would be *smaller* than \$1000 with a *larger* sample ($n > 554$), and *larger* with a *smaller* sample ($n < 554$).

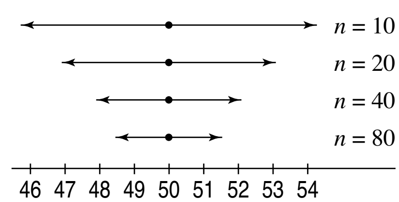
If all 1000 graduates respond, the margin of error would be $(1.96)(\$379.47) \doteq \$743.77$; if $n = 500$, the margin of error would be $(1.96)(\$536.66) \doteq \$1051.85$.

6.9. The (useful) response rate is $\frac{249}{5800} \doteq 0.0429$, or about 4.3%. The reported margin of error is probably unreliable because we know nothing about the 95.7% of students that did *not* provide (useful) responses; they may be more (or less) likely to charge education-related expenses.

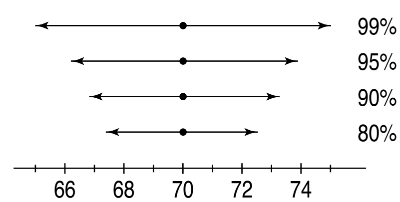
6.10. **(a)** The 95% confidence interval is $87 \pm 10 = 77$ to 97. (The sample size is not needed.) **(b)** Greater than 10: A wider margin of error is needed in order to be more confident that the interval includes the true mean.

Note: *If this result is based on a Normal distribution, the margin of error for 99% confidence would be roughly 13.1, because we multiply by 2.576 rather than 1.96.*

6.11. The margins of error are $1.96 \times 7/\sqrt{n}$, which yields 4.3386, 3.0679, 2.1693, and 1.5339. (And, of course, all intervals are centered at 50.) Interval width decreases with increasing sample size.



6.12. The margins of error are $z^* \times 14/\sqrt{49} = 2z^*$. With z^* equal to 1.282, 1.645, 1.960, and 2.576, this yields 2.564, 3.290, 3.920, and 5.152. (And, of course, all intervals are centered at 70.) Increasing confidence makes the interval wider.



6.13. **(a)** She did not divide the standard deviation by $\sqrt{n} = 20$. **(b)** Confidence intervals concern the population mean, not the sample mean. (The value of the sample mean is known to be 8.6; it is the population mean that we do not know.) **(c)** 95% is a confidence level, not a probability. Furthermore, it does not make sense to make probability statements about the population mean μ, which is an unknown constant (rather than a random quantity). **(d)** The large sample size does not affect the distribution of individual alumni ratings (the population distribution). The use of a Normal distribution is justified because the distribution *of the sample mean* is approximately Normal when the sample is large.

Note: *For part (c), a Bayesian statistician might view the population mean μ as a random quantity, but the viewpoint taken in the text is non-Bayesian.*

6.14. **(a)** The standard deviation should be divided by $\sqrt{100} = 10$, not by 100. **(b)** The correct interpretation is that (with 95% confidence) the *average* time spent at the site is between 3.71 and 4.69 hours. That is, the confidence interval is a statement about the population mean, not about the individual members. **(c)** To halve the margin of error, the sample size needs to be quadrupled, to about 400. (In fact, $n = 385$ would be enough.)

6.15. (a) To estimate the mean importance of recreation to college satisfaction, the 95% confidence interval for μ is

$$7.5 \pm 1.96\left(\frac{3.9}{\sqrt{2673}}\right) = 7.5 \pm 0.1478 = 7.3522 \text{ to } 7.6478$$

(b) The 99% confidence interval for μ is

$$7.5 \pm 2.576\left(\frac{3.9}{\sqrt{2673}}\right) = 7.5 \pm 0.1943 = 7.3057 \text{ to } 7.6943$$

6.16. We must assume that the 2673 students were chosen as an SRS (or something like it). The non-Normality of the population distribution is not a problem; we have a very large sample, so the central limit theorem applies.

6.17. For mean TRAP level, the margin of error is 2.29 U/l and the 95% confidence interval for μ is

$$13.2 \pm 1.96\left(\frac{6.5}{\sqrt{31}}\right) = 13.2 \pm 2.29 = 10.91 \text{ to } 15.49 \text{ U/l}$$

6.18. For mean OC level, the 95% confidence interval for μ is

$$33.4 \pm 1.96\left(\frac{19.6}{\sqrt{31}}\right) = 33.4 \pm 6.90 = 26.50 \text{ to } 40.30 \text{ ng/ml}$$

6.19. Scenario B has a smaller margin of error. Both samples would have the same value of z^* (1.96), but the value of σ would be smaller for (B) because we would have less variability in textbook cost for students in a single major.

Note: *Of course, at some schools, taking a sample of 100 sophomores in a given major is not possible. However, even if we sampled students from a number of institutions, we still might expect less variability within a given major than from a broader cross-section.*

6.20. We assume that the confidence interval was constructed using the methods of this chapter; that is, we assume that $\bar{x} \pm 1.96\sigma/\sqrt{2500}$. Then the center of the given confidence interval is $\bar{x} = \frac{1}{2}(\$45{,}330 + \$46{,}156) = \$45{,}743$, the margin of error for is $\$46{,}156 - \bar{x} = \$413$, and the 99% confidence margin of error is therefore $\$413 \cdot \frac{2.576}{1.96} = \$542.8$. Then the desired confidence interval is

$$(\$48{,}633 - \bar{x}) \pm \$542.8 = \$2347.2 \text{ to } \$3432.8;$$

that is, the average starting salary at this institution is about \$2300 to \$3400 less than that NACE mean.

6.21. (a) "The probability is about 0.95 that $\bar{x}$ is within 14 kcal/day of ... μ" (because 14 is two standard deviations). **(b)** This is simply another way of understanding the statement from part (a): If $|\bar{x} - \mu|$ is less than 14 kcal/day 95% of the time, then "about 95% of all samples will capture the true mean ... in the interval $\bar{x}$ plus or minus 14 kcal/day."

6.22. For the mean monthly rent for unfurnished one-bedroom apartments in Dallas, the 95% confidence interval for μ is

$$\$980 \pm 1.96\left(\frac{\$290}{\sqrt{10}}\right) = \$980 \pm \$179.74 = \$800.26 \text{ to } \$1159.74$$

6.23. No; This is a range of values for the mean rent, not for individual rents.

Note: *To find a range to include 95% of all rents, we should take $\mu \pm 2\sigma$ (or more precisely, $\mu \pm 1.96\sigma$), where μ is the (unknown) mean rent for all apartments, and σ is the standard deviation for all apartments (assumed to be \$290 in Exercise 6.22). If μ were equal to \$1050, for example, this range would be about \$470 to \$1630. However, because we do not actually know μ, we estimate it using $\bar{x}$, and to account for the variability in $\bar{x}$, we must widen the margin of error by a factor of $\sqrt{1+\frac{1}{n}}$. The formula $\bar{x} \pm 2\sigma\sqrt{1+\frac{1}{10}}$ is called a* prediction interval *for future observations. (Usually, such intervals are constructed with the t distribution, discussed in the Chapter 7, but the idea is the same.)*

6.24. If the distribution were roughly Normal, the 68–95–99.7 rule says that 68% of all measurements should be in the range 13.8 to 53.0 ng/ml, 95% should be between -5.8 and 72.6 ng/ml, and 99.7% should be between -25.4 and 92.2 ng/ml. Because the measurements cannot be negative, this suggests that the distribution must be skewed to the right. The Normal confidence interval should be fairly accurate nonetheless because the central limit theorem says that the distribution of the sample mean $\bar{x}$ will be roughly Normal.

6.25. **(a)** For the mean number of hours spent on the Internet, the 95% confidence interval for μ is

$$19 \pm 1.96\left(\frac{5.5}{\sqrt{1200}}\right) = 19 \pm 0.3112 = 18.6888 \text{ to } 19.3112 \text{ hours}$$

(b) No; this is a range of values for the mean time spent, not for individual times. (See also the comment in the solution to Exercise 6.23.)

6.26. **(a)** To change from hours to minutes, multiply by 60: $\bar{x}_m = 60\bar{x}_h = 1140$ and $\sigma_m = 60\sigma_h = 330$ minutes. **(b)** For mean time in minutes, the 95% confidence interval for μ is

$$1140 \pm 1.96\left(\frac{330}{\sqrt{1200}}\right) = 1140 \pm 18.67 = 1121.33 \text{ to } 1158.67 \text{ minutes}$$

(c) This interval can be found by multiplying the previous interval (18.6888 to 19.3112 hours) by 60.

6.27. **(a)** We can be 95% confident, but not *certain*. **(b)** We obtained the interval 85% to 95% by a method that gives a correct result (that is, includes the true mean) 95% of the time. **(c)** For 95% confidence, the margin of error is about two standard deviations (that is, $z^* = 1.96$), so $\sigma_{\text{estimate}} \doteq 2.5\%$. **(d)** No; confidence intervals only account for random sampling error.

6.28. **(a)** The standard deviation of the mean is $\sigma_{\bar{x}} = \frac{3.5}{\sqrt{20}} \doteq 0.7826$ mpg. **(b)** A stemplot (right) does not suggest any severe deviations from Normality. The mean of the 20 numbers in the sample is $\bar{x} = 43.17$ mpg. **(c)** If μ is the population mean fuel efficiency, the 95% confidence interval for μ is

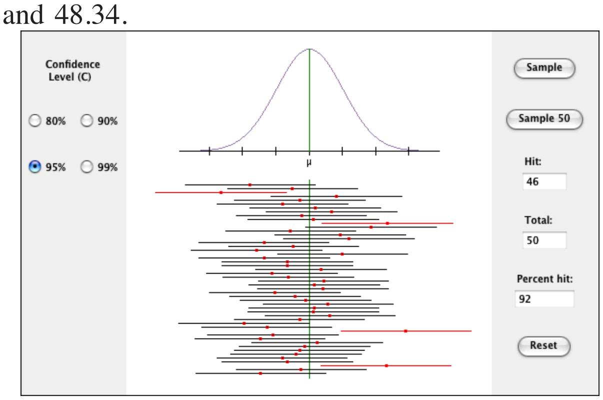
$$43.17 \pm 1.96\left(\frac{3.5}{\sqrt{20}}\right) = 43.17 \pm 1.5339 = 41.6361 \text{ to } 44.7039 \text{ mpg}$$

3	4
3	677
3	9
4	1
4	23333
4	445
4	667
4	88
5	0

6.29. Multiply by $\frac{1.609 \text{ km}}{1 \text{ mile}} \cdot \frac{1 \text{ gallon}}{3.785 \text{ liters}} \doteq 0.4251 \frac{\text{kpl}}{\text{mpg}}$. This gives $\bar{x}_{\text{kpl}} = 0.4251\bar{x}_{\text{mpg}} \doteq 18.3515$ and margin of error $1.96(0.4251\sigma_{\text{mpg}})/\sqrt{20} \doteq 0.6521$ kpl, so the 95% confidence interval is 17.6994 to 19.0036 kpl.

6.30. One sample screen is shown below, along with a sample stemplot of results. The number of hits will vary, but the distribution should follow a binomial distribution with $n = 50$ and $p = 0.95$, so we expect the average number of hits to be about 47.5. We also find that about 99.7% of individual counts should be 43 or more, and the mean hit count for 30 samples should be approximately Normal with mean 47.5 and standard deviation 0.2814—so almost all sample means should be between 46.66 and 48.34.

44	00
45	0000
46	00
47	00000000
48	000000
49	000
50	00000



6.31. $n = \left(\frac{(1.96)(6.5)}{1.5}\right)^2 \doteq 72.14$—take $n = 73$.

6.32. If we start with a sample of size k and lose 20% of the sample, we will end with $0.8k$. Therefore, we need to *increase* the sample size by 25%—that is, start with a sample of size $k = 1.25n$—so that we end with $(0.8)(1.25n) = n$. With $n = 73$, that means we should initially sample $k = 91.25$ (use 92) subjects.

6.33. No: Because the numbers are based on voluntary response rather than an SRS, the confidence interval methods of this chapter cannot be used; the interval does not apply to the whole population.

6.34. **(a)** For the mean of all repeated measurements, the 98% confidence interval for μ is

$$10.0023 \pm 2.326\left(\frac{0.0002}{\sqrt{5}}\right) = 10.0023 \pm 0.0002 = 10.0021 \text{ to } 10.0025 \text{ g}$$

(b) $n = \left(\frac{(2.326)(0.0002)}{0.0001}\right)^2 \doteq 21.64$—take $n = 22$.

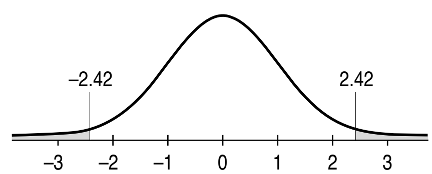
6.35. The number of hits has a binomial distribution with parameters $n = 5$ and $p = 0.95$, so the number of misses is binomial with $n = 5$ and $p = 0.05$. We can therefore use Table C to answer these questions. **(a)** The probability that all cover their means is $0.95^5 \doteq 0.7738$. (Or use Table C to find the probability of 0 misses.) **(b)** The probability that at least four intervals cover their means is $0.95^5 + 5(0.05)(0.95^4) \doteq 0.9774$. (Or use Table C to find the probability of 0 or 1 misses.)

6.36. The new design can be considered an improvement if the mean response μ to the survey is greater than 4. The null hypothesis should be H_0: $\mu = 4$; the alternative hypothesis could be either $\mu > 4$ or $\mu < 4$. The first choice would be appropriate if we want the default assumption (H_0) to be that the new design is *not* an improvement; that is, we will only conclude that the new design is better if we see compelling evidence to that effect. Choosing H_a: $\mu < 4$ would mean that the default assumption is that the new design is at least as good as the old one, and we will stick with that belief unless we see compelling evidence that it is worse.

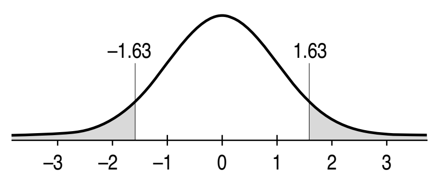
Students who are just learning about stating hypotheses might have difficulty choosing the alternative for this problem. In fact, either one is defensible, although the typical choice in such cases would be $\mu > 4$—that is, we give the benefit of the doubt to the old design and need convincing evidence that the new design is better.

6.37. If μ is the mean DXA reading for the phantom, we test H_0: $\mu = 1.4$ g/cm^2 versus H_a: $\mu \neq 1.4$ g/cm^2.

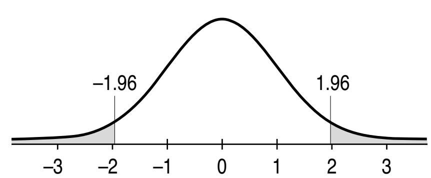
6.38. $P(Z > 2.42) = 0.0078$, so the two-sided P-value is $2(0.0078) = 0.0156$.



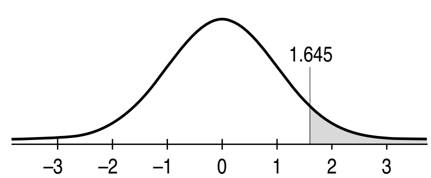
6.39. $P(Z < -1.63) = 0.0516$, so the two-sided P-value is $2(0.0516) = 0.1032$.



6.40. **(a)** For $P = 0.05$, the value of z is ± 1.96. **(b)** For a two-sided alternative, z is statistically significant at $\alpha = 0.05$ if $|z| > 1.96$.



6.41. **(a)** For $P = 0.05$, the value of z is 1.645. **(b)** For a one-sided alternative (on the positive side), z is statistically significant at $\alpha = 0.05$ if $z > 1.645$.



6.42. For $z^* = 2$ the P-value would be $2P(Z > 2) = 0.0456$, and for $z^* = 3$ the P-value would be $2P(Z > 3) = 0.0026$.

Note: *In other words, the Supreme Court uses α no bigger than about 0.05.*

6.43. **(a)** $z = \frac{27-25}{5/\sqrt{36}} = 2.4$. **(b)** For a one-sided alternative, $P = P(Z > 2.4) = 0.0082$. **(c)** For a two-sided alternative, double the one-sided P-value: $P = 0.0164$.

6.44. The test statistic is $z = \frac{0.514-0.5}{0.314/\sqrt{100}} \doteq 0.45$. This gives very little reason to doubt the null hypothesis ($\mu = 0.5$); in fact, the two-sided P-value is $P \doteq 0.6527$.

6.45. Recall the statement from the text: "A level α two-sided significance test rejects ... H_0: $\mu = \mu_0$ exactly when the value μ_0 falls outside a level $1-\alpha$ confidence interval for μ." **(a)** No; 30 is not in the 95% confidence interval because $P = 0.033$ means that we would reject H_0 at $\alpha = 0.05$. **(b)** Yes; 30 is in the 99% confidence interval because we would *not* reject H_0 at $\alpha = 0.01$.

6.46. See the quote from the text in the previous solution. **(a)** No, we would not reject $\mu = 58$ because 58 falls well inside the confidence interval, so the P-value is (much) greater than 0.05. **(b)** Yes, we would reject $\mu = 63$; the fact that 63 falls outside the 95% confidence interval means that $P < 0.05$.

Note: *The given confidence interval suggests that $\bar{x} = 57.5$, and if the interval was constructed using the Normal distribution, the standard error of the mean is about 2.25—half the margin of error. (The standard error might be less if it was constructed with a t distribution rather than the Normal distribution.) Then $\bar{x}$ is about 0.22 standard errors below $\mu = 58$—yielding $P \doteq 0.82$—and $\bar{x}$ is about 2.44 standard errors below $\mu = 63$, so that $P \doteq 0.015$.*

6.47. **(a)** Yes, we reject H_0 at $\alpha = 0.05$. **(b)** No, we do not reject H_0 at $\alpha = 0.01$. **(c)** We have $P = 0.039$; we reject H_0 at significance level α if $P < \alpha$.

6.48. **(a)** No, we do not reject H_0 at $\alpha = 0.05$. **(b)** No, we do not reject H_0 at $\alpha = 0.01$. **(c)** We have $P = 0.062$; we reject H_0 at significance level α if $P < \alpha$.

6.49. **(a)** One of the one-sided P-values is half as big as the two-sided P-value (0.022); the other is $1 - 0.022 = 0.978$. **(b)** Suppose the null hypothesis is H_0: $\mu = \mu_0$. The smaller P-value (0.022) goes with the one-sided alternative that is consistent with the observed data; for example, if $\bar{x} > \mu_0$, then $P = 0.022$ for the alternative $\mu > \mu_0$.

6.50. **(a)** The null hypothesis should be a statement about μ, not $\bar{x}$. **(b)** The standard deviation of the sample mean is $5/\sqrt{30}$. **(c)** $\bar{x} = 45$ would not make us inclined to believe that $\mu > 50$ over the (presumed) null hypothesis $\mu = 50$. **(d)** Even if we fail to reject H_0, we are not sure that it is true.

Note: *That is, "not rejecting H_0" is different from "knowing that H_0 is true." This is the same distinction we make about a jury's verdict in a criminal trial: If the jury finds the defendant "not guilty," that does* not *necessarily mean that they are sure he/she is innocent. It simply means that they were not sufficiently convinced of his/her guilt.*

6.51. **(a)** Hypotheses should be stated in terms of the population mean, not the sample mean. **(b)** The null hypothesis H_0 should be that there is no change ($\mu = 21.2$). **(c)** A small P-value is needed for significance; $P = 0.98$ gives no reason to reject H_0. **(d)** We compare the P-value, not the z-statistic, to α. (In this case, such a small value of z would have a very large P-value—close to 0.5 for a one-sided alternative, or close to 1 for a two-sided alternative.)

6.52. **(a)** We are checking to see if the proportion p increased, so we test H_0: $p = 0.88$ versus H_a: $p > 0.88$. **(b)** The professor believes that the mean μ for the morning class will be higher, so we test H_0: $\mu = 75$ versus H_a: $\mu > 75$. **(c)** Let μ be the mean response (for the population of all students who read the newspaper). We are trying to determine if students are neutral about the change, or if they have an opinion about it, with no preconceived idea about the direction of that opinion, so we test H_0: $\mu = 0$ versus H_a: $\mu \neq 0$.

6.53. **(a)** If μ is the mean score for the population of placement-test students, then we test H_0: $\mu = 77$ versus H_a: $\mu \neq 77$ because we have no prior belief about whether placement-test students will do better or worse. **(b)** If μ is the mean time to complete the maze with rap music playing, then we test H_0: $\mu = 20$ seconds versus H_a: $\mu > 20$ seconds because we believe rap music will make the mice finish more slowly. **(c)** If μ is the mean area of the apartments, we test H_0: $\mu = 880$ ft^2 versus H_a: $\mu < 880$ ft^2, because we suspect the apartments are smaller than advertised.

6.54. **(a)** If p_m and p_f are the proportions of (respectively) males and females who like MTV best, we test H_0: $p_m = p_f$ versus H_a: $p_m > p_f$. **(b)** If μ_A and μ_B are the mean test scores for each group, we test H_0: $\mu_A = \mu_B$ versus H_a: $\mu_A > \mu_B$. **(c)** If ρ is the (population) correlation between time spent at social network sites and self-esteem, we test H_0: $\rho = 0$ versus H_a: $\rho < 0$.

Note: *In each case, the parameters identified refer to the respective* populations, *not the samples.*

6.55. **(a)** H_0: $\mu = \$42{,}800$ versus H_a: $\mu > \$42{,}800$, where μ is the mean household income of mall shoppers. **(b)** H_0: $\mu = 0.4$ hr versus H_a: $\mu \neq 0.4$ hr, where μ is this year's mean response time.

6.56. **(a)** For H_a: $\mu > \mu_0$, the P-value is $P(Z > 1.63) = 0.0516$.
(b) For H_a: $\mu < \mu_0$, the P-value is $P(Z < 1.63) = 0.9484$.
(c) For H_a: $\mu \neq \mu_0$, the P-value is $2P(Z > 1.63) = 2(0.0516) = 0.1032$.

6.57. **(a)** For H_a: $\mu > \mu_0$, the P-value is $P(Z > -1.82) = 0.9656$.
(b) For H_a: $\mu < \mu_0$, the P-value is $P(Z < -1.82) = 0.0344$.
(c) For H_a: $\mu \neq \mu_0$, the P-value is $2P(Z < -1.82) = 2(0.0344) = 0.0688$.

6.58. Recall the statement from the text: "A level α two-sided significance test rejects ... H_0: $\mu = \mu_0$ exactly when the value μ_0 falls outside a level $1 - \alpha$ confidence interval for μ." **(a)** No, 30 is not in the 95% confidence interval because $P = 0.032$ means that we would reject H_0 at $\alpha = 0.05$. **(b)** No, 30 is not in the 90% confidence interval because we would also reject H_0 at $\alpha = 0.10$.

6.59. See the quote from the text in the previous solution. **(b)** Yes, we would reject H_0: $\mu = 24$; the fact that 24 falls outside the 90% confidence interval means that $P < 0.10$. **(a)** No, we would not reject H_0: $\mu = 30$ because 30 falls inside the confidence interval, so $P > 0.10$.

Note: *The given confidence interval suggests that $\bar{x} = 28.5$, and if the interval was constructed using the Normal distribution, the standard error of the mean is about 1.75—half the margin of error. (The standard error might be less if it was constructed with a t distribution rather than the Normal distribution.) Then $\bar{x}$ is about 2.57 standard errors above $\mu = 24$—yielding $P \doteq 0.01$—and $\bar{x}$ is about 0.86 standard errors below $\mu = 30$, so that $P \doteq 0.39$.*

6.60. The study presumably examined malarial infection rates in two groups of subjects—one with bed nets and one without. The observed differences between the two groups were so large that they would be unlikely to occur by chance if bed nets had no effect. Specifically, if the groups were the same, and we took many samples, the difference in malarial infections would be so large less than 0.1% of the time.

6.61. $P = 0.09$ means there is some evidence for the wage decrease, but it is not significant at the $\alpha = 0.05$ level. Specifically, the researchers observed that average wages for peer-driven students were 13% lower than average wages for ability-driven students, but (when considering overall variation in wages) such a difference might arise by chance 9% of the time, even if student motivation had no effect on wages.

6.62. If the presence of pig skulls were not an indication of wealth, then differences similar to those observed in this study would occur less than 1% of the time by chance.

6.63. Even if the two groups (the health and safety class, and the statistics class) had the same level of alcohol awareness, there might be some difference in our sample due to chance. The difference observed was large enough that it would rarely arise by chance. The reason for this difference might be that health issues related to alcohol use are probably discussed in the health and safety class.

6.64. Even if scores had not changed over time, random fluctuation might cause the mean in 2009 to be different from the 2007 mean. However, in this case the difference was so great that it is unlikely to have occurred by chance; specifically, such a difference would arise less than 5% of the time if the actual mean had not changed. We therefore conclude that the mean did change from 2007 to 2009.

6.65. If μ is the mean difference between the two groups of children, we test H_0: $\mu = 0$ versus H_a: $\mu \neq 0$. The test statistic is $z = \frac{4-0}{1.2} \doteq 3.33$, for which software reports $P \doteq 0.0009$—very strong evidence against the null hypothesis.

Note: *The exercise reports the standard deviation of the mean, rather than the sample standard deviation; that is, the reported value has already been divided by $\sqrt{238}$.*

6.66. If μ is the mean north-south location, we wish to test H_0: $\mu = 100$ versus H_a: $\mu \neq 100$. We find $z = \frac{99.74 - 100}{58/\sqrt{584}} \doteq -0.11$; this is not significant—$P = 2(0.4562) = 0.9124$—so we have no reason to doubt a uniform distribution based on this test.

6.67. If μ is the mean east-west location, the hypotheses are H_0: $\mu = 100$ versus H_a: $\mu \neq 100$ (as in the previous exercise). For testing these hypotheses, we find $z = \frac{113.8 - 100}{58/\sqrt{584}} \doteq 5.75$. This is highly significant ($P < 0.0001$), so we conclude that the trees are not uniformly spread from east to west.

6.68. For testing these hypotheses, we find $z = \frac{10.2 - 8.9}{2.5/\sqrt{6}} \doteq 1.27$. This is not significant ($P = 0.1020$); there is not enough evidence to conclude that these sonnets were not written by our poet. (That is, we cannot reject H_0.)

6.69. **(a)** $z = \frac{127.8 - 115}{30/\sqrt{25}} \doteq 2.13$, so the P-value is $P = P(Z > 2.13) = 0.0166$. This is strong evidence that the older students have a higher SSHA mean. **(b)** The important assumption is that this is an SRS from the population of older students. We also assume a Normal distribution, but this is not crucial provided there are no outliers and little skewness.

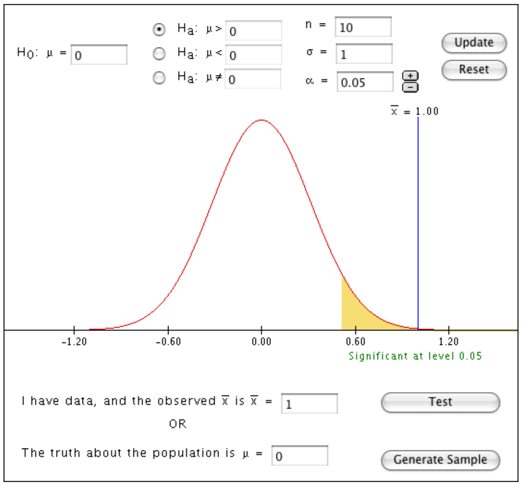
6.70. **(a)** Because we suspect that athletes might be deficient, we use a one-sided alternative: H_0: $\mu = 2811.5$ kcal/day versus H_a: $\mu < 2811.5$ kcal/day. **(b)** The test statistic is $z = \frac{2403.7 - 2811.5}{880/\sqrt{201}} \doteq -6.57$, for which $P < 0.0001$. There is strong evidence of below-recommended caloric consumption among female Canadian high-performance athletes.

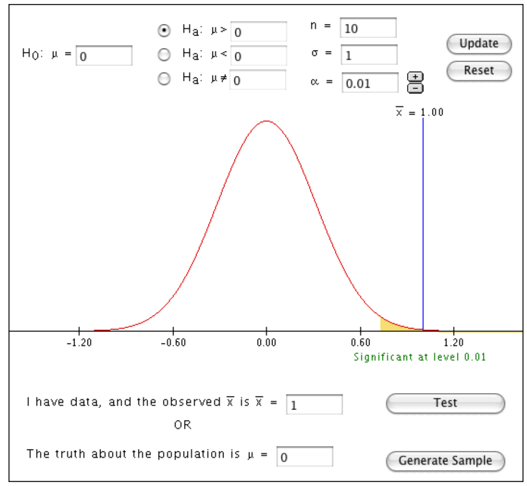
6.71. **(a)** H_0: $\mu = 0$ mpg versus H_a: $\mu \neq 0$ mpg, where μ is the mean difference. **(b)** The mean of the 20 differences is $\bar{x} = 2.73$, so $z = \frac{2.73 - 0}{3/\sqrt{20}} \doteq 4.07$, for which $P < 0.0001$. We conclude that $\mu \neq 0$ mpg; that is, we have strong evidence that the computer's reported fuel efficiency differs from the driver's computed values.

6.72. A debt of \$3817 in the West would be equivalent to a debt of (\$3817)(1.06) $\doteq$ \$4046 in the Midwest, for a difference of \$4046 − \$3260 = \$786. With the hypotheses given in Example 6.10, and the standard deviation (\$374) from Example 6.11, the test statistic is $z = \frac{\$786 - \$0}{\$374} \doteq 2.10$. The P-value is $P = 2P(Z \geq 2.10) \doteq 0.0358$; recall that, for the unadjusted data, the P-value was 0.1362. Adjusting for the differing value of a dollar strengthens the evidence against H_0, enough that it is now significant at the 5% level.

6.73. For (b) and (c), either compare with the critical values in Table D or determine the P-value (0.0336). **(a)** H_0: $\mu = 0.9$ mg versus H_a: $\mu > 0.9$ mg. **(b)** Yes, because $z > 1.645$ (or because $P < 0.05$). **(c)** No, because $z < 2.326$ (or because $P > 0.01$).

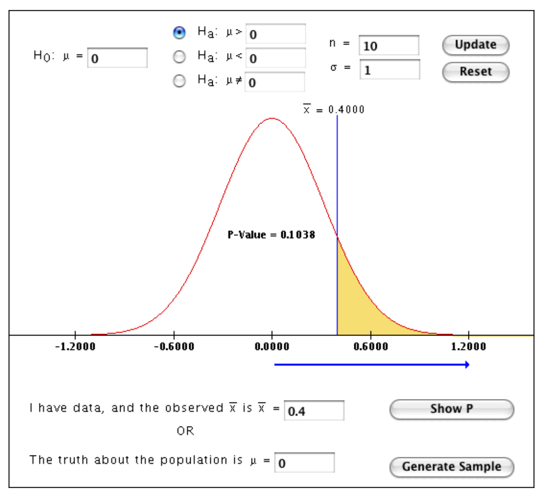
6.74. A sample screen (for $\bar{x} = 1$) is shown below on the left. As one can judge from the shading under the Normal curve, $\bar{x} = 0.5$ is not significant, but 0.6 is. (In fact, the cutoff is about 0.52, which is approximately $1.645/\sqrt{10}$.)





6.75. See the sample screen (for $\bar{x} = 1$) above on the right. As one can judge from the shading under the Normal curve, $\bar{x} = 0.7$ is not significant, but 0.8 is. (In fact, the cutoff is about 0.7354, which is approximately $2.326/\sqrt{10}$.) Smaller α means that $\bar{x}$ must be farther away from μ_0 in order to reject H_0.

6.76. A sample screen (for $\bar{x} = 0.4$) is shown on the right. The P-values given by the applet are listed in the table below; as $\bar{x}$ moves farther away from μ_0, P decreases.



$\bar{x}$	P	$\bar{x}$	P
0.1	0.3745	0.6	0.0287
0.2	0.2643	0.7	0.0136
0.3	0.1711	0.8	0.0057
0.4	0.1038	0.9	0.0022
0.5	0.0571	1	0.0008

6.77. When a test is significant at the 5% level, it means that if the null hypothesis were true, outcomes similar to those seen are expected to occur fewer than 5 times in 100 repetitions of the experiment or sampling. "Significant at the 10% level" means we have observed something that occurs in fewer than 10 out of 100 repetitions (when H_0 is true). Something that occurs "fewer than 5 times in 100 repetitions" also occurs "fewer than 10 times in 100 repetitions," so significance at the 5% level implies significance at the 10% level (or any higher level).

6.78. Something that occurs "fewer than 5 times in 100 repetitions" is not necessarily as rare as something that occurs "less than once in 100 repetitions," so a test that is significant at 5% is not necessarily significant at 1%.

6.79. Using Table D or software, we find that the 0.005 critical value is 2.576, and the 0.0025 critical value is 2.807. Therefore, if $2.576 < |z| < 2.807$—that is, either $2.576 < z < 2.807$ or $-2.807 < z < -2.576$—then z would be significant at the 1% level, but not at the 0.5% level.

6.80. As $2.326 < 2.52 < 2.576$, the two-sided P-value is between $2(0.005) = 0.01$ and $2(0.01) = 0.02$. (Software tells us that $P \doteq 0.012$, consistent with the observation that z is close to 2.576.)

6.81. As $0.63 < 0.674$, the one-sided P-value is $P > 0.25$. (Software gives $P = 0.2643$.)

6.82. Because $1.645 < 1.92 < 1.960$, the P-value is between $2(0.025) = 0.05$ and $2(0.05) = 0.10$. From Table A, $P = 2(0.0274) = 0.0548$.

6.83. Because the alternative is two-sided, the answer for $z = -1.92$ is the same as for $z = 1.92$: $-1.645 > -1.92 > -1.960$, so Table D says that $0.05 < P < 0.10$, and Table A gives $P = 2(0.0274) = 0.0548$.

6.84. **(a)** $z = \frac{541.4 - 525}{100/\sqrt{100}} = 1.64$. This is not significant at $\alpha = 0.05$ because $z < 1.645$ (or $P = 0.0505$). **(b)** $z = \frac{541.5 - 525}{100/\sqrt{100}} = 1.65$. This *is* significant at $\alpha = 0.05$ because $z > 1.645$ (or $P = 0.0495$). **(c)** Fixed-level significance tests require that we draw a line between "significant" and "not significant"; in this example, we see evidence on each side of that line. The 5% significance level is a guideline, not a sacred edict. P-values are more informative ways to convey the strength of the evidence.

6.85. In order to determine the effectiveness of alarm systems, we need to know the percent of all homes with alarm systems, and the percent of burglarized homes with alarm systems. For example, if only 10% of all home have alarm systems, then we should compare the proportion of burglarized homes with alarm systems to 10%, not 50%.

An alternate (but rather impractical) method would be to sample homes and classify them according to whether or not they had an alarm system, and also by whether or not they had experienced a break-in at some point in the recent past. This would likely require a very large sample in order to get a sufficiently large count of homes that had experienced break-ins.

6.86. Finding something to be "statistically significant" is not really useful unless the significance level is sufficiently small. While there is some freedom to decide what "sufficiently small" means, $\alpha = 0.20$ would lead the student to incorrectly reject H_0 one-fifth of the time, so it is clearly a bad choice.

6.87. The first test was barely significant at $\alpha = 0.05$, while the second was significant at any reasonable α.

6.88. One can learn something from negative results; for example, a study that finds no benefit from a particular treatment is at least useful in terms of what will *not* work. Furthermore, reviewing such results might point researchers to possible future areas of study.

6.89. A significance test answers only Question b. The P-value states how likely the observed effect (or a stronger one) is if H_0 is true, and chance alone accounts for deviations from what we expect. The observed effect may be significant (very unlikely to be due to chance) and yet not be of practical importance. And the calculation leading to significance *assumes* a properly designed study.

6.90. Based on the description, this seems to have been an experiment (not just an observational study), so a statistically significant outcome suggests that vitamin C is effective in preventing colds.

6.91. (a) If SES had no effect on LSAT results, there would still be some difference in scores due to chance variation. "Statistically insignificant" means that the observed difference was no more than we might expect from that chance variation. **(b)** If the results are based on a small sample, then even if the null hypothesis were not true, the test might not be sensitive enough to detect the effect. Knowing the effects were small tells us that the statistically insignificant test result did not occur merely because of a small sample size.

6.92. These questions are addressed in the summary for Section 6.3. **(a)** Failing to reject H_0 does not mean that H_0 is true. **(b)** This is correct; a difference that is statistically significant might not be practically important. (This does not mean that these are opposites; a difference *could* be both statistically and practically significant.) **(c)** This might be technically true, but in order for the analysis to be meaningful, the data must satisfy the assumptions of the analysis. **(d)** Searching for patterns and then testing their significance can lead to false positives (that is, we might reject the null hypothesis incorrectly). If a pattern is observed, we should collect new data to test if it is present.

6.93. In each case, we find the test statistic z by dividing the observed difference ($2453.7 - 2403.7 = 50$ kcal/day) by $880/\sqrt{n}$. **(a)** For $n = 100$, $z \doteq 0.57$, so $P = P(Z > 0.57) = 0.2843$. **(b)** For $n = 500$, $z \doteq 1.27$, so $P = P(Z > 1.27) = 0.1020$. **(c)** For $n = 2500$, $z \doteq 2.84$, so $P = P(Z > 2.84) = 0.0023$.

6.94. The study may have rejected $\mu = \mu_0$ (or some other null hypothesis), but with such a large sample size, such a rejection might occur even if the actual mean (or other parameter) differs only slightly from μ_0. For example, there might be no practical importance to the difference between $\mu = 10$ and $\mu = 10.5$.

6.95. We expect more variation with small sample sizes, so even a large difference between $\bar{x}$ and μ_0 (or whatever measures are appropriate in our hypothesis test) might not turn out to be significant. If we were to repeat the test with a larger sample, the decrease in the standard error might give us a small enough P-value to reject H_0.

6.98. When many variables are examined, "significant" results will show up by chance, so we should not take it for granted that the variables identified are really indicative of future success. In order to decide if they are appropriate, we should track this year's trainees and compare the success of those from urban/suburban backgrounds with the rest, and likewise compare those with a degree in a technical field with the rest.

6.100. We expect 50 tests to be statistically significant: Each of the 1000 tests has a 5% chance of being significant, so the number of significant tests has a binomial distribution with $n = 1000$ and $p = 0.05$, for which the mean is $np = 50$.

6.101. $P = 0.00001 = \frac{1}{100{,}000}$, so we would need $n = 100{,}000$ tests in order to expect one P-value of this size (assuming that all null hypotheses are true). That is why we reject H_0 when we see P-values such as this: It indicates that our results would rarely happen if H_0 were true.

6.102. Using $\alpha/6 \doteq 0.008333$ as the cutoff, the fourth ($P = 0.003$) and sixth ($P < 0.001$) tests are significant.

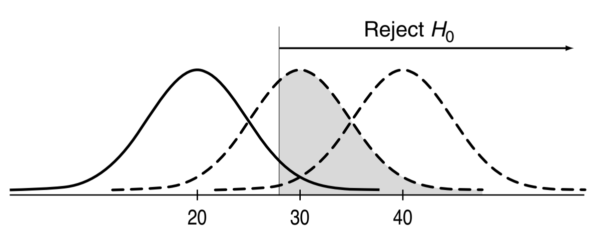
6.103. Using $\alpha/12 \doteq 0.004167$ as the cutoff, we reject the fifth ($P = 0.002$) and eleventh ($P < 0.002$) tests.

6.104. The power of this study is far lower than what is generally desired—for example, it is well below the "80% standard" mentioned in the text. For the specified effect, 35% power means that, if the effect is present, we will only detect it 35% of the time. With such a small chance of detecting an important difference, the study should probably not be run (unless the sample size is increased to give sufficiently high power).

6.105. A larger sample gives more information and therefore gives a better chance of detecting a given alternative; that is, larger samples give more power.

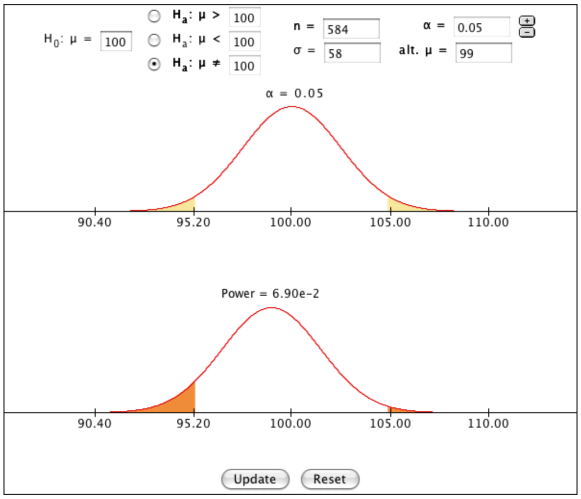
6.106. The power for $\mu = -3$ is 0.82—the same as the power for $\mu = 3$—because both alternatives are an equal distance from the null value of μ. (The symmetry of two-sided tests with the Normal distribution means that we only need to consider the size of the difference, not the direction.)

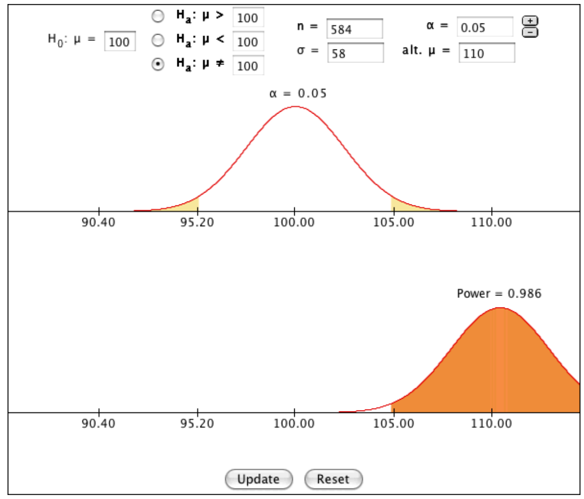
6.107. The power for $\mu = 40$ will be higher than 0.6, because larger differences are easier to detect. The picture on the right shows one way to illustrate this (assuming Normal distributions): The solid curve (centered at 20) is the distribution under the null hypothesis, and the two dashed curves represent the alternatives $\mu = 30$ and $\mu = 40$. The shaded region under the middle curve is the power against $\mu = 30$; that is, that shaded region is 60% of the area under that curve. The power against $\mu = 40$ would be the corresponding area under the rightmost curve, which would clearly be greater than 0.6.



6.108. The applet finds that the power is approximately 0.069.

6.109. The applet reports the power as 0.986.





6.110. (a) For the alternative H_a: $\mu > 168$, we reject H_0 at the 5% significance level if $z > 1.645$. **(b)** $\frac{\bar{x}-168}{27/\sqrt{70}} > 1.645$ when $\bar{x} > 168 + 1.645 \cdot \frac{27}{\sqrt{70}} \doteq 173.31$. **(c)** When $\mu = 173$, the probability of rejecting H_0 is

$$P(\bar{x} > 173.31) = P\left(\frac{\bar{x}-173}{27/\sqrt{70}} > \frac{173.31-173}{27/\sqrt{70}}\right) \doteq P(Z > 0.10) = 0.4602.$$

(d) The power of this test is not up to the 80% standard suggested in the text; he should collect a larger sample.

Note: *Software gives a slightly different answer for the power in part (c), but the conclusion in part (d) is the same. To achieve 80% power against $\mu = 173$, we need $n = 180$.*

6.111. We reject H_0 when $z > 2.326$, which is equivalent to $\bar{x} > 450 + 2.326 \cdot \frac{100}{\sqrt{500}} \doteq 460.4$, so the power against $\mu = 460$ is

$$\begin{aligned} P(\text{reject } H_0 \text{ when } \mu = 460) &= P(\bar{x} > 460.4 \text{ when } \mu = 460) \\ &= P\left(Z > \frac{460.4-460}{100/\sqrt{500}}\right) \doteq P(Z > 0.09) = 0.4641. \end{aligned}$$

This is quite a bit less than the "80% power" standard.

6.112. (a) $P(\text{Type I error}) = P(X \le 2$ when the distribution is $p_0) = 0.4$.
(b) $P(\text{Type II error}) = P(X > 2$ when the distribution is $p_1) = 0.4$.

6.113. (a) The hypotheses are "subject should go to college" and "subject should join work force." The two types of errors are recommending someone go to college when (s)he is better suited for the work force, and recommending the work force for someone who should go to college. **(b)** In significance testing, we typically wish to decrease the probability of wrongly rejecting H_0 (that is, we want α to be small); the answer to this question depends on which hypothesis is viewed as H_0.

Note: *For part (a), there is no clear choice for which should be the null hypothesis. In the past, when fewer people went to college, one might have chosen "work force" as H_0—that is, one might have said, "we'll assume this student will join the work force unless we are convinced otherwise." Presently, roughly two-thirds of graduates attend college, which might suggest H_0 should be "college."*

6.114. This is probably not a confidence interval; it is not intended to give an estimate of the mean income, but rather it gives the range of incomes earned by all (or most) telemarketers working for this company.

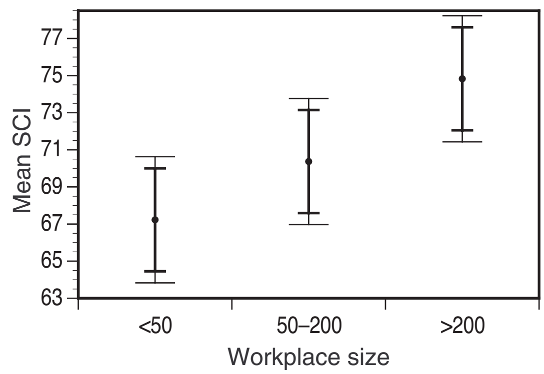
6.115. From the description, we might surmise that we had two (or more) groups of students—say, an exercise group and a control (or no-exercise) group. **(a)** For example, if μ is the mean difference in scores between the two groups, we might test H_0: $\mu = 0$ versus H_a: $\mu \neq 0$. (Assuming we had no prior suspicion about the effect of exercise, the alternative should be two-sided.) **(b)** With $P = 0.38$, we would not reject H_0. In plain language: The results observed do not differ greatly from what we would expect if exercise had no effect on exam scores. **(c)** For example: Was this an experiment? What was the design? How big were the samples?

6.116. (a) For each proportion, the estimated standard deviation (in the column labeled "SD") is $\sqrt{\hat{p}(1-\hat{p})/n}$; the margin of error ("m.e.") is $1.96 \times$ SD. **(b)** The first two groups (professional and managerial) had the highest stress levels, followed by the next two (administrative and sales), then the next three (mechanical, service, and operator). Stress levels were lowest for farm workers. **(c)** Because data was compiled over two years, some individuals might have been included more than once, which violates the independence assumption of the binomial distribution.

Occupation	$\hat{p}$	SD	m.e.	Conf. interval
Professional	0.23	0.00851	0.01667	0.2133 to 0.2467
Managerial	0.22	0.00820	0.01607	0.2039 to 0.2361
Administrative	0.17	0.00782	0.01532	0.1547 to 0.1853
Sales	0.15	0.00839	0.01645	0.1336 to 0.1664
Mechanical	0.12	0.00730	0.01432	0.1057 to 0.1343
Service	0.13	0.00661	0.01295	0.1171 to 0.1429
Operator	0.12	0.00616	0.01208	0.1079 to 0.1321
Farm	0.08	0.01135	0.02225	0.0577 to 0.1023

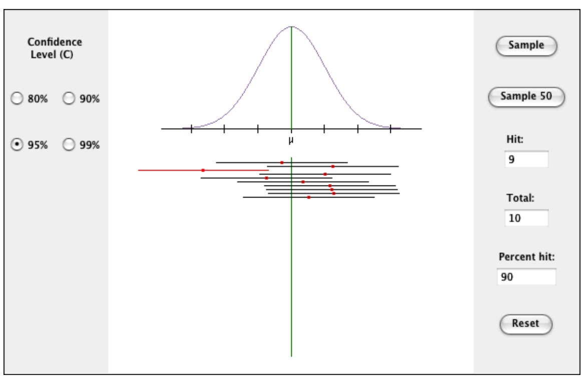
6.117. (a) Because all standard deviations and sample sizes are the same, the margin of error for all intervals is $1.96 \times 19/\sqrt{180} \doteq 2.7757$. The confidence intervals are listed in the table on the following page. **(b)** The plot on the following page shows the error bars for the confidence intervals of part (a), and also for part (c). The limits for (a) are the thicker lines which do not extend as far above and below the mean. **(c)** With $z^* = 2.40$, the margin of error for all intervals is $2.40 \times 19/\sqrt{180} \doteq 3.3988$. The confidence intervals are listed in the table below and are shown in the plot (the thinner lines with the wider dashes). **(d)** When we use $z^* = 2.40$ to adjust for the fact that we are making three "simultaneous" confidence intervals, the margin of error is larger, so the intervals overlap more.

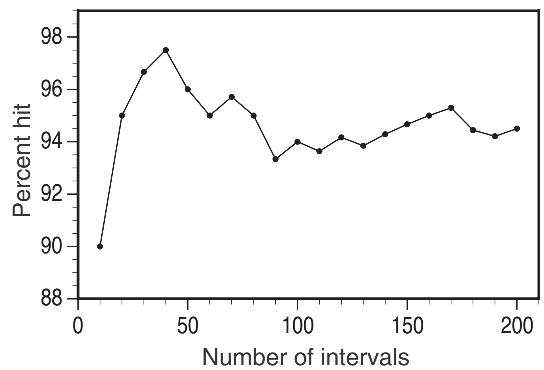
Workplace size	Mean SCI
< 50	64.45 to 70.01
50–200	67.59 to 73.15
> 200	72.05 to 77.61
< 50	63.83 to 70.63
50–200	66.97 to 73.77
> 200	71.43 to 78.23



6.118. Shown below is a sample screenshot from the applet and an example of what the resulting plot might look like. Most students (99.7% of them) should find that their final proportion is between 0.90 and 1; 90% will have a proportion between 0.925 and 0.975.

Note: *For each n (number of intervals), the number of "hits" would have a binomial distribution with p = 0.95, but these counts would not be independent; for example, if we knew there were 28 hits after 30 tries, we would know that there could be no more than 38 after 40 tries.*

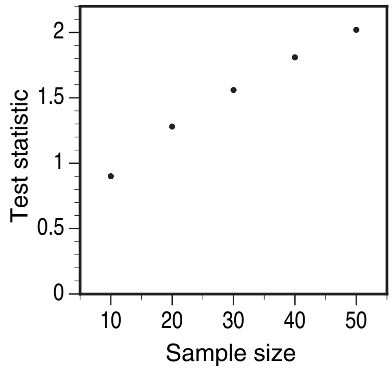


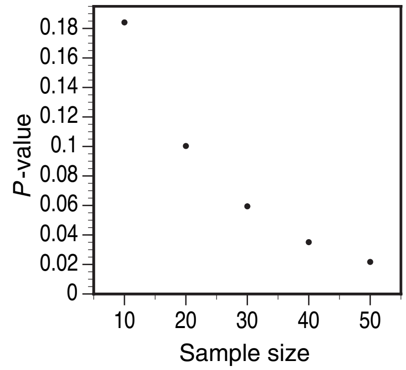


6.119. A sample screenshot and example plot are not shown but would be similar to those shown above for the previous exercise. Most students (99.4% of them) should find that their final proportion is between 0.84 and 0.96; 85% will have a proportion between 0.87 and 0.93.

6.120. For $n = 10$, $z = \frac{1-0}{14/\sqrt{10}} \doteq 0.90$, for which $P = 0.1841$. For the other sample sizes, the computations are similar; the resulting table and graphs are shown on the following page. We see that sample size increases the value of the test statistic (assuming the mean is the same), which in turn decreases the size of the P-value.

n	z	P
10	0.90	0.1841
20	1.28	0.1003
30	1.56	0.0594
40	1.81	0.0351
50	2.02	0.0217





6.121. (a) $\bar{x} = 5.\overline{3}$ mg/dl, so $\bar{x} \pm 1.960\sigma/\sqrt{6}$ is 4.6132 to 6.0534 mg/dl. **(b)** To test H_0: $\mu = 4.8$ mg/dl versus H_a: $\mu > 4.8$ mg/dl, we compute $z = \frac{\bar{x}-4.8}{0.9/\sqrt{6}} \doteq 1.45$ and $P \doteq 0.0735$. This is not strong enough to reject H_0.

Note: *The confidence interval in (a) would allow us to say without further computation that, against a two-sided alternative, we would have P > 0.05. Because we have a one-sided alternative, we could conclude from the confidence interval that P > 0.025, but that is not enough information to draw a conclusion.*

6.122. (a) The 90% confidence interval for μ is

$$145 \pm 1.645\left(\frac{8}{\sqrt{15}}\right) = 145 \pm 3.3979 = 141.6021 \text{ to } 148.3979 \text{ mg/g}$$

(b) Our hypotheses are H_0: $\mu = 140$ mg/g versus H_a: $\mu > 140$ mg/g. The test statistic is $z = \frac{145-140}{8/\sqrt{15}} = 2.42$, so the P-value is $P = P(Z > 2.42) = 0.0078$. This is strong evidence against H_0; we conclude that the mean cellulose content is higher than 140 mg/g. **(c)** We must assume that the 15 cuttings in our sample are an SRS. Because our sample is not too large, the population should be Normally distributed, or at least not extremely nonnormal.

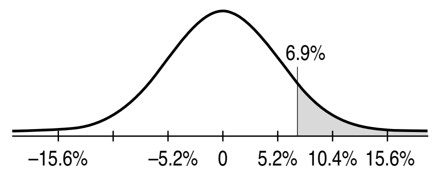
6.123. (a) The stemplot is reasonably symmetric for such a small sample. **(b)** $\bar{x} = 30.4$ μg/l; $30.4 \pm (1.96)(7/\sqrt{10})$ gives 26.0614 to 34.7386 μg/l. **(c)** We test H_0: $\mu = 25$ μg/l versus H_a: $\mu > 25$ μg/l. $z = \frac{30.4-25}{7/\sqrt{10}} \doteq 2.44$, so $P = 0.0073$. (We knew from (b) that it had to be smaller than 0.025). This is fairly strong evidence against H_0; the beginners' mean threshold is higher than 25 μg/l.

2	034
2	
3	01124
3	6
4	3

6.124. (a) The intended population is probably "the American public"; the population that was actually sampled was "citizens of Indianapolis (with listed phone numbers)." **(b)** Take $\bar{x} \pm 1.96s/\sqrt{201}$; these intervals are listed on the right. **(c)** The confidence intervals do not overlap at all; in particular, the *lower* confidence limit of the rating for pharmacies is higher than the *upper* confidence limit for the other stores. This indicates that the pharmacies are *really* rated higher.

Food stores	15.22 to 22.12
Mass merchandisers	27.77 to 36.99
Pharmacies	43.68 to 53.52

6.125. **(a)** Under H_0, $\bar{x}$ has a $N(0\%, 55\%/\sqrt{104}) \doteq N(0\%, 5.3932\%)$ distribution. **(b)** $z = \frac{6.9-0}{55/\sqrt{104}} \doteq 1.28$, so $P = P(Z > 1.28) = 0.1003$. **(c)** This is not significant at $\alpha = 0.05$. The study gives *some* evidence of increased compensation, but it is not very strong; similar results would happen about 10% of the time just by chance.



6.126. No: "Significant at $\alpha = 0.01$" *does* mean that the null hypothesis is unlikely, but only in the sense that the evidence (from the sample) would not occur very often if H_0 were true. There is no probability associated with H_0; it is either true or it is not.

Note: *Bayesian statistics views the parameter we wish to estimate as having a probability distribution; with that viewpoint, it would make sense to speak of "the probability that H_0 is true." This textbook does not take the Bayesian approach.*

6.127. Yes. That's the heart of why we care about statistical significance. Significance tests allow us to discriminate between random differences ("chance variation") that might occur when the null hypothesis is true, and differences that are unlikely to occur when H_0 is true.

6.128. If p is the probability that red occurs, we test H_0: $p = \frac{18}{38}$ versus H_a: $p \neq \frac{18}{38}$.

6.129. For each sample, find $\bar{x}$, then take $\bar{x} \pm 1.96(4/\sqrt{12}) = \bar{x} \pm 2.2632$.

We "expect" to see that 95 of the 100 intervals will include 25 (the true value of μ); binomial computations show that (about 99% of the time) 90 or more of the 100 intervals will include 20.

6.130. For each sample, find $\bar{x}$, then compute $z = \frac{\bar{x}-25}{4/\sqrt{12}}$. Choose a significance level α and the appropriate cutoff point—for example, with $\alpha = 0.10$, reject H_0 if $|z| > 1.645$; with $\alpha = 0.05$, reject H_0 if $|z| > 1.96$.

If, for example, $\alpha = 0.05$, then we "expect" to reject H_0 (that is, make the wrong decision) only 5 of the 100 times.

6.131. For each sample, find $\bar{x}$, then compute $z = \frac{\bar{x}-23}{4/\sqrt{12}}$. Choose a significance level α and the appropriate cutoff point (z^*)—for example, with $\alpha = 0.10$, reject H_0 if $|z| > 1.645$; with $\alpha = 0.05$, reject H_0 if $|z| > 1.96$.

Because the true mean is 25, $Z = \frac{\bar{x}-25}{4/\sqrt{12}}$ has a $N(0, 1)$ distribution, so the probability that we will accept H_0 is $P\left(-z^* < \frac{\bar{x}-23}{4/\sqrt{12}} < z^*\right) = P(-z^* < Z + 1.7321 < z^*) = P(-1.7321 - z^* < Z < -1.7321 + z^*)$. If $\alpha = 0.10$ ($z^* = 1.645$), this probability is $P(-3.38 < Z < -0.09) = 0.4637$; if $\alpha = 0.05$ ($z^* = 1.96$), this probability is $P(-3.69 < Z < 0.23) = 0.5909$. For smaller α, the probability will be larger. Thus we "expect" to (wrongly) accept H_0 about half the time (or more), and correctly reject H_0 about half the time or less. (The probability of rejecting H_0 is essentially the power of the test against the alternative $\mu = 25$.)

6.132. The test statistics and P-values are in the table below, computed as described in the text; for example, for conscientiousness, $z = \frac{3.80-3.88}{0.10} = -0.8$. Because we are performing 13 tests, we should use the Bonferroni procedure (see Exercise 6.102) or some other multiple-test method. For an overall significance level of α, Bonferroni requires individual-test significance at $\alpha/13$; for $\alpha = 0.05$, this means we need $P < 0.0038$. Therefore, the only significant differences (for $\alpha = 0.05$ or $\alpha = 0.01$) are for handshake strength and grip.

Characteristic	z	P
Conscientiousness	−0.8	0.4238
Extraversion	0.9	0.3682
Agreeableness	−2.2	0.0278
Emotional stability	1.5	0.1336
Openness to experience	0.5	0.6170
Overall handshake	2.3	0.0214
Handshake strength	5.3	< 0.0001*
Handshake vigor	1.7	0.0892
Handshake grip	3.8	0.0002*
Handshake duration	1.5	0.1336
Eye contact	−0.6	0.5486
Professional dress	−2.0	0.0456
Interviewer assessment	−0.58	0.5626

Chapter 7 Solutions

7.1. **(a)** The standard error of the mean is $\frac{s}{\sqrt{n}} = \frac{\$96}{\sqrt{16}} = \$24$. **(b)** The degrees of freedom are df $= n - 1 = 15$.

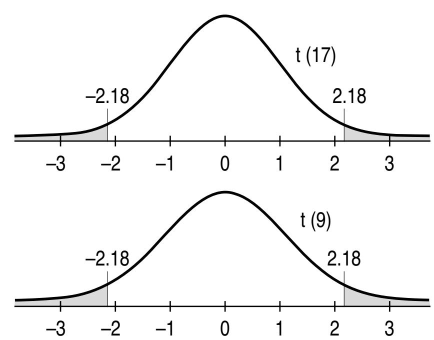
7.2. In each case, use df $= n - 1$; if that number is not in Table D, drop to the *lower* degrees of freedom. **(a)** For 95% confidence and df = 10, use $t^* = 2.228$. **(b)** For 99% confidence and df = 32, we drop to df = 30 and use $t^* = 2.750$. (Software gives $t^* = 2.7385$ for df = 32.) **(c)** For 90% confidence and df = 249, we drop to df = 100 and use $t^* = 1.660$. (Software gives $t^* = 1.6510$ for df = 249.)

7.3. For the mean monthly rent, the 95% confidence interval for μ is

$$\$613 \pm 2.131\left(\frac{\$96}{\sqrt{16}}\right) = \$613 \pm \$51.14 = \$561.86 \text{ to } \$664.14$$

7.4. The margin of error for 90% confidence would be smaller (so the interval would be narrower) because we are taking a greater risk—specifically, a 10% risk—that the interval does *not* include the true mean μ.

7.5. **(a)** Yes, $t = 2.18$ is significant when $n = 18$. This can be determined either by comparing to the df = 17 line in Table D (where we see that $t > 2.110$, the 2.5% critical value) or by computing the two-sided P-value (which is $P = 0.0436$). **(b)** No, $t = 2.18$ is not significant when $n = 10$, as can be seen by comparing to the df = 9 line in Table D (where we see that $t < 2.262$, the 2.5% critical value) or by computing the two-sided P-value (which is $P = 0.0572$). **(c)** Student sketches will likely be indistinguishable from Normal distributions; careful students may try to show that the $t(9)$ distribution is shorter in the center and heavier to the left and right ("in the tails") than the $t(17)$ distribution (as is the case here), but in reality, the difference is nearly imperceptible.



7.6. For the hypotheses H_0: $\mu = \$550$ versus H_a: $\mu > \$550$, we find $t = \frac{613 - 550}{96/\sqrt{16}} = 2.625$ with df = 15, for which $P \doteq 0.0096$. We have strong evidence against H_0, and conclude that the mean rent is greater than \$550.

7.7. Software will typically give a more accurate value for t^* than that given in Table D, and will not round off intermediate values such as the standard deviation. Otherwise, the details of this computation are the same as what is shown in the textbook: df = 7, $t^* = 2.3646$, $6.75 \pm t^*(3.8822/\sqrt{8}) = 6.75 \pm 3.2456 = 3.5044$ to 9.9956, or about 3.5 to 10.0 hours per month.

7.8. We wish to test H_0: $\mu = 0$ versus H_a: $\mu \neq 0$, where μ is the mean of (drink A rating) minus (drink B rating). (We could also subtract in the other direction.) Compute the differences for each subject (-2, 5, 2, 10, and 7), then find the mean and standard deviation of these differences: $\bar{x} = 4.4$ and $s \doteq 4.6152$. Therefore, $t = \frac{4.4-0}{4.6152/\sqrt{5}} \doteq 2.1318$ with df = 4, for which $P \doteq 0.1000$. We do not have enough evidence to conclude that there is a difference in preference.

7.9. About -1.33 to 10.13: Using the mean and standard deviation from the previous exercise, the 95% confidence interval for μ is

$$4.4 \pm 2.7765\left(\frac{4.6152}{\sqrt{5}}\right) = 4.4 \pm 5.7305 = -1.3305 \text{ to } 10.1305$$

(This is the interval produced by software; using the critical value $t^* = 2.776$ from Table D gives -1.3296 to 10.1296.)

7.10. See also the solutions to Exercises 1.36, 1.74, and 1.150. The CO_2 data are sharply right-skewed (clearly non-Normal). However, the robustness of the t procedures should make them safe for this situation because the sample size is large ($n = 48$). The bigger question is whether we can treat the data as an SRS; we have recorded CO_2 emissions for every country with a population over 20 million, rather than a random sample.

7.11. The distribution is clearly non-Normal, but the sample size ($n = 63$) should be sufficient to overcome this, especially in the absence of strong skewness. One might question the independence of the observations; it seems likely that after 40 or so tickets had been posted for sale, that someone listing a ticket would look at those already posted for an idea of what price to charge.

If we were to use t procedures, we would presumably take the viewpoint that these 63 observations come from a larger population of hypothetical tickets for this game, and we are trying to estimate the mean μ of that population. However, because (based on the histogram in Figure 1.33) the population distribution is likely bimodal, the mean μ might not be the most useful summary of a bimodal distribution.

7.12. The power would be greater because larger differences (like $\mu > 1$) are easier to detect.

7.13. As was found in Example 7.9, we reject H_0 if $t = \frac{\bar{x}}{1.5/\sqrt{20}} \geq 1.729$, which is equivalent to $\bar{x} \geq 0.580$. The power we seek is $P(\bar{x} \geq 0.580$ when $\mu = 1.1)$, which is:

$$P\left(\frac{\bar{x} - 1.1}{1.5/\sqrt{20}} \geq \frac{0.580 - 1.1}{1.5/\sqrt{20}}\right) = P(Z \geq -1.55) = 0.9394$$

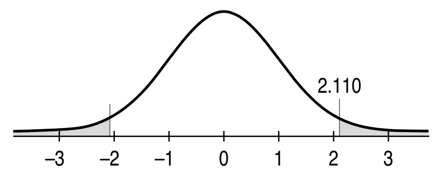
7.14. We test H_0: median $= 0$ versus H_a: median $\neq 0$—or equivalently, H_0: $p = 1/2$ versus H_a: $p \neq 1/2$, where p is the probability that the rating for drink A is higher. We note that four of the five differences are positive (and none are 0). The P-value is $2P(X \geq 4) = 0.375$ from a $B(5, 0.5)$ distribution; there is not enough evidence to conclude that the median ratings are different.

Minitab output: Sign test of median = 0 versus median ≠ 0

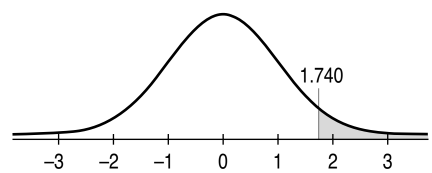
	N	BELOW	EQUAL	ABOVE	P-VALUE	MEDIAN
Diff	5	1	0	4	0.3750	5.000

7.15. **(a)** df $= 10$, $t^* = 2.228$. **(b)** df $= 21$, $t^* = 2.080$. **(c)** df $= 21$, $t^* = 1.721$. **(d)** For a given confidence level, t^* (and therefore the margin of error) decreases with increasing sample size. For a given sample size, t^* increases with increasing confidence.

7.16. This t distribution has df $= 17$. The 2.5% critical value is 2.110, so we reject H_0 when $t < -2.110$ or $t > 2.110$.

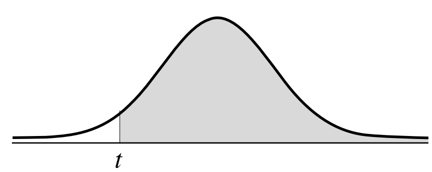


7.17. The 5% critical value for a t distribution with df $= 17$ is 1.740. Only one of the one-sided options (reject H_0 when $t > 1.740$) is shown; the other is simply the mirror image of this sketch (shade the area to the left of -1.740, and reject when $t < -1.740$).



7.18. Because the value of $\bar{x}$ is positive, which supports the direction of the alternative ($\mu > 0$), the P-value for the one-sided test is half as big as that for the two-sided test: $P = 0.037$.

7.19. $\bar{x} = -15.3$ would support the alternative $\mu < 0$, and for that alternative, the P-value would still be 0.037. For the alternative $\mu > 0$ given in Exercise 7.18, the P-value is 0.963. Note that in the sketch shown, no scale has been given, because in the absence of a sample size, we do not know the degrees of freedom. Nevertheless, the P-value for the alternative $\mu > 0$ is the area above the computed value of the test statistic t, which will be the opposite of that found when $\bar{x} = 15.3$. As the area below t is 0.037, the area above this point must be 0.963.



7.20. **(a)** df $= 15$. **(b)** $1.753 < t < 2.131$. **(c)** $0.025 < P < 0.05$. **(d)** $t = 2.10$ is significant at 5%, but not at 1%. **(e)** From software, $P \doteq 0.0265$.

7.21. **(a)** df $= 27$. **(b)** $1.703 < t < 2.052$. **(c)** Because the alternative is two-sided, we double the upper-tail probabilities to find the P-value: $0.05 < P < 0.10$. **(d)** $t = 2.01$ is not significant at either level (5% or 1%). **(e)** From software, $P \doteq 0.0546$.

7.22. **(a)** df $= 13$. **(b)** Because $2.282 < |t| < 2.650$, the P-value is between $0.01 < P < 0.02$. **(c)** From software, $P \doteq 0.0121$.

7.23. Let P be the given (two-sided) P-value, and suppose that the alternative is $\mu > \mu_0$. If $\bar{x}$ is greater than μ_0, this supports the alternative over H_0. However, if $\bar{x} < \mu_0$, we would not take this as evidence against H_0 because $\bar{x}$ is on the "wrong" side of μ_0. So, if the value of $\bar{x}$ is on the "correct" side of μ_0, the one-sided P-value is simply $P/2$. However, if the value of $\bar{x}$ is on the "wrong" side of μ_0, the one-sided P-value is $1 - P/2$ (which will always be at least 0.5, so it will never indicate significant evidence against H_0).

7.24. **(a)** The distribution is slightly right-skewed, and the largest observation stands out from the rest (although it does not quite qualify as an outlier using the $1.5 \times IQR$ rule). **(b)** The reasonably large sample should be sufficient to overcome the mild non-Normality of the data, and because it was based on a random sample from a large population, t procedures should be appropriate. **(c)** $\bar{x} \doteq 119.0667$ and $s \doteq 29.5669$ friends, so the standard error is $s/\sqrt{30} \doteq 5.3982$. The critical value for 95% confidence is $t^* = 2.045$, so the margin of error is 11.04. **(d)** With 95% confidence, the mean number of Facebook friends at this university is between 108.03 and 130.11.

7	24
8	355
9	679
10	3456
11	012899
12	0678
13	7
14	8
15	248
16	0
17	1
18	
19	3

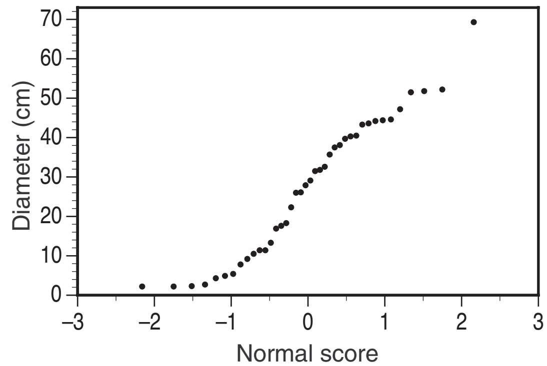
7.25. **(a)** If μ is the mean number of uses a person can produce in 5 minutes after witnessing rudeness, we wish to test H_0: $\mu = 10$ versus H_a: $\mu < 10$. **(b)** $t = \frac{7.88 - 10}{2.35/\sqrt{34}} \doteq -5.2603$, with df $= 33$, for which $P < 0.0001$. This is very strong evidence that witnessing rudeness decreases performance.

7.26. **(a)** A stemplot (shown) or a histogram shows no outliers and no particular skewness. (In fact, for such a small sample, it suggests no striking deviations from Normality.) The use of t methods seems to be safe. **(b)** The mean is $\bar{x} \doteq 43.17$ mpg, the standard deviation is $s \doteq 4.4149$ mpg, and the standard error is $s/\sqrt{20} \doteq 0.9872$ mpg. For df $= 19$, the 2.5% critical value is $t^* \doteq 2.093$, so the margin of error is $t^* s/\sqrt{20} \doteq 2.0662$ mpg. **(c)** The 95% confidence interval is 41.1038 to 45.2362 mpg.

3	4
3	677
3	9
4	1
4	23333
4	445
4	667
4	88
5	0

7.27. **(a)** A stemplot (right) reveals that the distribution has two peaks and a high value (not quite an outlier). Both the stemplot and quantile plot show that the distribution is not Normal. The five-number summary is 2.2, 10.95, 28.5, 41.9, 69.3 (all in cm); a boxplot is not shown, but the long "whisker" between Q_3 and the maximum is an indication of the skewness. **(b)** Maybe: We have a large enough sample to overcome the non-Normal distribution, but we are sampling from a small population. **(c)** The mean is $\bar{x} = 27.29$ cm, $s \doteq 17.7058$ cm, and the margin of error is $t^* \cdot s/\sqrt{40}$:

0	222244
0	579
1	0113
1	678
2	2
2	6679
3	112
3	5789
4	0033444
4	7
5	112
5	
6	
6	9



	df	t^*	Interval
Table D	30	2.042	$27.29 \pm 5.7167 = 21.57$ to 33.01 cm
Software	39	2.0227	$27.29 \pm 5.6626 = 21.63$ to 32.95 cm

(d) One could argue for either answer. We chose a random sample from this tract, so the main question is, can we view trees in this tract as being representative of trees elsewhere?

7.28. **(a)** We wish to test H_0: $\mu = 3421.7$ kcal/day versus H_a: $\mu < 3421.7$ kcal/day. **(b)** The test statistic is $t = \frac{3077 - 3421.7}{987/\sqrt{114}} \doteq -3.73$, with df $= 113$, for which $P \doteq 0.0002$. **(c)** Starting with the average shortfall $3421.7 - 3077.0 = 344.7$ kcal/day, the mean deficiency is (with 95% confidence) between about 160 and 530 kcal/day.

	df	t^*	Interval
Table D	100	1.984	$344.7 \pm 183.4030 = 161.2970$ to 528.1030
Software	113	1.9812	$344.7 \pm 183.1423 = 161.5577$ to 527.8423

7.29. **(a)** The distribution is not Normal—there were lots of 1s and 10s—but the nature of the scale means that there can be no extreme outliers, so with a sample of size 60, the t methods should be acceptable. **(b)** The mean is $\bar{x} \doteq 5.9$, $s \doteq 3.7719$, and the margin of error is $t^* \cdot s/\sqrt{60}$:

	df	t^*	Interval
Table D	50	2.009	$5.9 \pm 0.9783 = 4.9217$ to 6.8783
Software	59	2.0010	$5.9 \pm 0.9744 = 4.9256$ to 6.8744

1	0000000000000000
2	0000
3	0
4	0
5	00000
6	000
7	0
8	000000
9	00000
10	000000000000000000

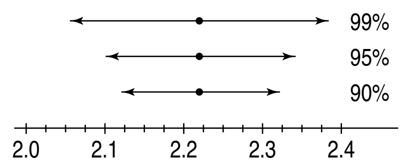
(c) Because this is not a random sample, it may not represent other children well.

7.30. (a) The distribution cannot be Normal because all values must be (presumably) integers between 0 and 4. **(b)** The sample size (282) should make the t methods appropriate because the distribution of ratings can have no outliers. **(c)** The margin of error is $t^* \cdot s/\sqrt{282}$, which is either 0.1611 (Table D) or 0.1591 (software):

	df	t^*	Interval
Table D	100	2.626	$2.22 \pm 0.1611 = 2.0589$ to 2.3811
Software	281	2.5934	$2.22 \pm 0.1591 = 2.0609$ to 2.3791

(d) The sample might not represent children from other locations well (or perhaps more accurately, it might not represent well the opinions of parents of children from other locations).

7.31. These intervals are constructed as in the previous exercise, except for the choice of t^*. We see that the width of the interval increases with confidence level.



		df	t^*	Interval
90% confidence	Table D	100	1.660	$2.22 \pm 0.1018 = 2.1182$ to 2.3218
	Software	281	1.6503	$2.22 \pm 0.1012 = 2.1188$ to 2.3212
95% confidence	Table D	100	1.984	$2.22 \pm 0.1217 = 2.0983$ to 2.3417
	Software	281	1.9684	$2.22 \pm 0.1207 = 2.0993$ to 2.3407

7.32. (a) For example, Subject 1's weight change is $61.7 - 55.7 = 6$ kg. **(b)** The mean change is $\bar{x} = 4.73125$ kg and the standard deviation is $s \doteq 1.7457$ kg. **(c)** $\text{SE}_{\bar{x}} = s/\sqrt{16} \doteq 0.4364$ kg; for df = 15, $t^* = 2.131$, so the margin of error for 95% confidence is ± 0.9300 (software: ± 0.9302). Based on a method that gives correct results 95% of the time, the mean weight change is 3.8012 to 5.6613 kg (software: 3.8010 to 5.6615 kg). **(d)** $\bar{x} = 10.40875$ lb, $s \doteq 3.8406$ lb, and the 95% confidence interval is 8.3626 to 12.4549 lb (software: 8.3622 to 12.4553 lb). **(e)** H_0 is $\mu = 16$ lb. The test statistic is $t \doteq -5.823$ with df = 15, which is highly significant evidence ($P < 0.0001$) against H_0 (unless H_a is $\mu > 16$ lb). **(f)** The data suggest that the excess calories were not converted into weight; the subjects must have used this energy some other way. (See the next exercise for more information.)

7.33. (a) $t = \frac{328 - 0}{256/\sqrt{16}} \doteq 5.1250$ with df = 15, for which $P \doteq 0.0012$. There is strong evidence of a change in NEAT. **(b)** With $t^* = 2.131$, the 95% confidence interval is 191.6 to 464.4 kcal/day. This tells us how much of the additional calories might have been burned by the increase in NEAT: It consumed 19% to 46% of the extra 1000 kcal/day.

7.34. (a) For the differences, $\bar{x} = \$112.5$ and $s \doteq \$123.7437$. **(b)** We wish to test H_0: $\mu = 0$ versus H_a: $\mu > 0$, where μ is the mean difference between Jocko's estimates and those of the other garage. (The alternative hypothesis is one-sided because the insurance adjusters suspect that Jocko's estimates may be too high.) For this test, we find $t = \frac{112.5 - 0}{123.7437/\sqrt{10}} \doteq 2.87$ with df = 9, for which $P = 0.0092$ (Minitab output below). This is significant evidence against H_0—that is, we have good reason to believe that Jocko's estimates are higher. **(c)** The 95% confidence interval with df $=$ 9 is $\bar{x} \pm 2.262\, s/\sqrt{10} = \$112.5 \pm \$88.5148 = \$23.99$ to $\$201.01$. (The software interval is \$23.98 to \$201.02.) **(d)** Student answers may vary; based on the confidence interval, one could justify any answer in the range \$25,000 to \$200,000.

Minitab output: Test of μ = 0 vs μ > 0

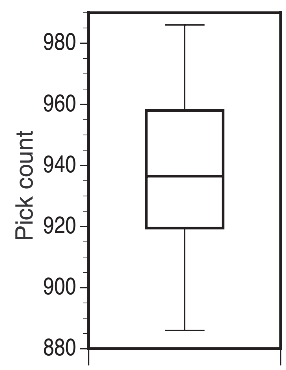
```
Variable      N      Mean     StDev   SE Mean        T    P-Value
Diff         10     112.5     123.7      39.1     2.87     0.0092
```

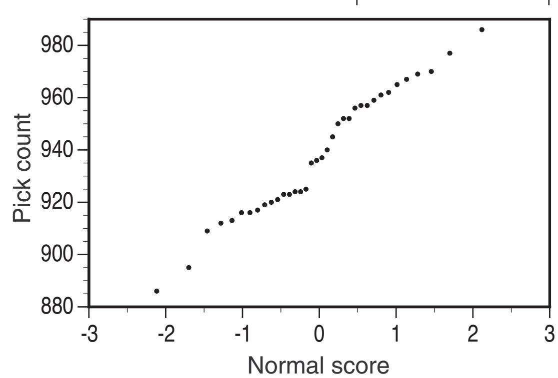
7.35. **(a)** We wish to test H_0: $\mu_c = \mu_d$ versus H_a: $\mu_c \neq \mu_d$, where μ_c is the mean computer-calculated mpg and μ_d is the mean mpg computed by the driver. Equivalently, we can state the hypotheses in terms of μ, the mean difference between computer- and driver-calculated mpgs, testing H_0: $\mu = 0$ versus H_a: $\mu \neq 0$. **(b)** With mean difference $\bar{x} \doteq 2.73$ and standard deviation $s \doteq 2.8015$, the test statistic is $t = \frac{2.73 - 0}{2.8015/\sqrt{20}} \doteq 4.3580$ with df $= 19$, for which $P \doteq 0.0003$. We have strong evidence that the results of the two computations are different.

7.36. **(a)** The plots are on the right; the five-number summary (in units of "picks") is

Min	Q_1	M	Q_3	Max
886	919.5	936.5	958	986

88	6
89	5
90	9
91	236679
92	0133445
93	567
94	05
95	0226779
96	12579
97	07
98	6





There are no outliers or particular skewness, but the stemplot reveals two peaks. (The boxplot gives no evidence of the two peaks; they are visible in the quantile plot, but it takes a fair amount of thought—or practice—to observe this in a quantile plot.) **(b)** While the distribution is non-Normal, there are no outliers or strong skewness, so the sample size $n = 36$ should make the t procedures reasonably safe. **(c)** The mean is $\bar{x} \doteq 938.\overline{2}$, the standard deviation is $s \doteq 24.2971$, and the standard error of the mean is $s/\sqrt{36} \doteq 4.0495$. (All are in units of picks.) **(d)** The 90% confidence interval for the mean number of picks in a 1-pound bag is:

	df	t^*	Interval
Table D	30	1.697	$938.\overline{2} \pm 6.8720 = 931.3502$ to 945.0943
Software	35	1.6896	$938.\overline{2} \pm 6.8420 = 931.3803$ to 945.0642

7.37. **(a)** To test H_0: $\mu = 925$ picks versus H_a: $\mu > 925$ picks, we have $t = \frac{938.\overline{2} - 925}{24.2971/\sqrt{36}} \doteq 3.27$ with df $= 35$, for which $P \doteq 0.0012$. **(b)** For H_0: $\mu = 935$ picks versus H_a: $\mu > 935$ picks, we have $t = \frac{938.\overline{2} - 935}{24.2971/\sqrt{36}} \doteq 0.80$, again with df $= 35$, for which $P \doteq 0.2158$. **(c)** The 90%

confidence interval from the previous exercise was 931.4 to 945.1 picks, which includes 935, but not 925. For a test of H_0: $\mu = \mu_0$ versus H_a: $\mu \neq \mu_0$, we know that $P < 0.10$ for values of μ_0 outside the interval, and $P > 0.10$ if μ_0 is inside the interval. The one-sided P-value would be half of the two-sided P-value.

7.38. The 90% confidence interval is $3.8 \pm t^*(1.02/\sqrt{1783})$. With Table D, take df = 1000 and $t^* = 1.646$; with software, take df = 1782 and $t^* = 1.6457$. Either way, the confidence interval is 3.7602 to 3.8398.

7.39. **(a)** The differences are spread from −0.018 to 0.020 g, with mean $\bar{x} = -0.0015$ and standard deviation $s \doteq 0.0122$ g. A stemplot is shown on the right; the sample is too small to make judgments about skewness or symmetry. **(b)** For H_0: $\mu = 0$ versus H_a: $\mu \neq 0$, we find $t = \frac{-0.0015-0}{s/\sqrt{8}} \doteq -0.347$ with df = 7, for which $P = 0.7388$. We cannot reject H_0 based on this sample. **(c)** The 95% confidence interval for μ is

$$-0.0015 \pm 2.365\left(\frac{0.0122}{\sqrt{8}}\right) = -0.0015 \pm 0.0102 = -0.0117 \text{ to } 0.0087 \text{ g}$$

−1	85
−1	
−0	65
−0	
0	2
0	55
1	
1	
2	0

(d) The subjects from this sample may be representative of future subjects, but the test results and confidence interval are suspect because this is not a random sample.

7.40. **(a)** The differences are spread from −31 to 45 g/cm^2, with mean $\bar{x} = 4.625$ and standard deviation $s \doteq 26.8485$ g/cm^2. A stemplot is shown on the right; the sample is too small to make judgments about skewness or symmetry. **(b)** For H_0: $\mu = 0$ versus H_a: $\mu \neq 0$, we find $t = \frac{4.625-0}{s/\sqrt{8}} \doteq 0.487$ with df = 7, for which $P = 0.6410$. We cannot reject H_0 based on this sample. **(c)** The 95% confidence interval for μ is

$$4.625 \pm 2.365\left(\frac{26.8485}{\sqrt{8}}\right) = 4.625 \pm 22.4494 = -17.8244 \text{ to } 27.0744 \text{ g/cm}^2$$

−3	1
−2	8
−1	
−0	4
0	16
1	3
2	
3	5
4	5

(d) See the answer to part (d) of the previous exercise.

7.41. **(a)** We test H_0: $\mu = 0$ versus H_a: $\mu > 0$, where μ is the mean change in score (that is, the mean improvement). **(b)** The distribution is slightly left-skewed, with mean $\bar{x} = 2.5$ and $s \doteq 2.8928$. **(c)** $t = \frac{2.5-0}{s/\sqrt{20}} \doteq 3.8649$, df = 19, and $P = 0.0005$; there is strong evidence of improvement in listening test scores. **(d)** With df = 19, we have $t^* = 2.093$, so the 95% confidence interval is 1.1461 to 3.8539.

−0	6
−0	
−0	
−0	0
0	0011
0	222333333
0	
0	66666

7.42. Graphical summaries may vary; shown on the right is a stemplot, after omitting the clear outlier (2631 seconds). The distribution is sharply right-skewed; in fact, based on the $1.5 \times IQR$ criterion, the top *eight* call lengths qualify as outliers. Given the sharp skewness of the data and the large number of outliers, the five-number summary is preferred, and we should be cautious about relying on the confidence interval. All the numerical summaries below are in units of seconds:

0	00000000111233444555555566677777778889
1	00001112222334444457789
2	00127789
3	2678
4	367
5	
6	
7	00
8	
9	5
10	
11	4

$\bar{x}$	s	Min	Q_1	M	Q_3	Max
196.575	342.0215	1	54.5	103.5	200	2631

The 95% confidence interval for the mean is roughly 120 to 273 seconds:

	df	t^*	Interval
Table D	70	1.994	$196.575 \pm 76.2489 = 120.3261$ to 272.8239 sec
Software	79	1.9905	$196.575 \pm 76.1132 = 120.4618$ to 272.6882 sec

7.43. The distribution is fairly symmetrical with no outliers. The mean IQ is $\bar{x} = 114.98\overline{3}$ and the standard deviation is $s \doteq 14.8009$. The 95% confidence interval is $\bar{x} \pm t^*(s/\sqrt{60})$, which is about 111.1 to 118.8:

	df	t^*	Interval
Table D	50	2.009	111.1446 to 118.8221
Software	59	2.0010	111.1598 to 118.8068

Because all students in the sample came from the same school, this *might* adequately describe the mean IQ at this school, but the sample could not be considered representative of all fifth graders.

8	12
8	9
9	04
9	67
10	01112223
10	568999
11	0002233444
11	5677788
12	223444
12	56778
13	01344
13	6799
14	2
14	5

7.44. We have data for all countries with population at least 20 million, so this cannot be considered a random sample of (say) all countries.

7.45. We test H_0: median $= 0$ versus H_a: median > 0—or equivalently, H_0: $p = 1/2$ versus H_a: $p > 1/2$, where p is the probability that Jocko's estimate is higher. One difference is 0; of the nine non-zero differences, seven are positive. The P-value is $P(X \geq 7) = 0.0898$ from a $B(9, 0.5)$ distribution; there is not quite enough evidence to conclude that Jocko's estimates are higher. In Exercise 7.34 we were able to reject H_0; here we cannot.

Note: *The failure to reject H_0 in this case is because with the sign test, we pay attention only to the* sign *of each difference, not the* size. *In particular, the negative differences are each given the same "weight" as each positive difference, in spite of the fact that the negative differences are only −\$50 and −\$75, while most of the positive differences are larger. See the "Caution" about the sign test on page 425 of the text.*

Minitab output: Sign test of median = 0 versus median > 0

```
                N  BELOW  EQUAL  ABOVE   P-VALUE     MEDIAN
Diff           10      2      1      7    0.0898      125.0
```

7.46. We test H_0: median $= 0$ versus H_a: median $\neq 0$. The Minitab output below gives $P = 1$ because there were four positive and four negative differences, giving us no reason to doubt H_0. (This is the same conclusion we reached with the t test, for which $P = 0.7388$.)

Minitab output: Sign test of median = 0 versus median ≠ 0

```
                N  BELOW  EQUAL  ABOVE   P-VALUE     MEDIAN
opdiff          8      4      0      4    1.0000   -0.00150
```

7.47. We test H_0: median $= 0$ versus H_a: median $\neq 0$. There were three negative and five positive differences, so the P-value is $2P(X \geq 5)$ for a binomial distribution with parameters $n = 8$ and $p = 0.5$. From Table C or software (Minitab output below), we have $P = 0.7266$, which gives no reason to doubt H_0. The t test P-value was 0.6410.

Minitab output: Sign test of median = 0 versus median ≠ 0

```
                N  BELOW  EQUAL  ABOVE   P-VALUE     MEDIAN
opdiff          8      3      0      5    0.7266      3.500
```

7.48. We test H_0: median $= 0$ versus H_a: median > 0, or H_0: $p = 1/2$ versus H_a: $p > 1/2$. Three of the 20 differences are zero; of the other 17, 16 are positive. The P-value is $P(X \geq 16)$ for a $B(17, 0.5)$ distribution. While Table C cannot give us the exact value of this probability, if we weaken the evidence by pretending that the three zero differences were negative and look at the $B(20, 0.5)$ distribution, we can estimate that $P < 0.0059$—enough information to reject the null hypothesis. In fact, software reports the P-value as 0.0001. (For the t test, we found $P = 0.0005$.)

Minitab output: Sign test of median = 0 versus median > 0

```
                N  BELOW  EQUAL  ABOVE   P-VALUE     MEDIAN
gain           20      1      3     16    0.0001      3.000
```

7.49. We test H_0: median $= 0$ versus H_a: median > 0, or H_0: $p = 1/2$ versus H_a: $p > 1/2$. Out of the 20 differences, 17 are positive (and none equal 0). The P-value is $P(X \geq 17)$ for a $B(20, 0.5)$ distribution. From Table C or software (Minitab output below), we have $P = 0.0013$, so we reject H_0 and conclude that the results of the two computations are different. (Using a t test, we found $P \doteq 0.0003$, which led to the same conclusion.)

Minitab output: Sign test of median = 0 versus median > 0

```
                N  BELOW  EQUAL  ABOVE   P-VALUE     MEDIAN
diff           20      3      0     17    0.0013      3.000
```

7.50. After taking logarithms, the 90% confidence interval is $\bar{x} \pm t^*(s/\sqrt{5})$. For df $= 4$, $t^* = 2.132$, and the confidence intervals are as shown in the table. (As we would expect, after exponentiating to undo the logarithms, both intervals are equivalent except for rounding differences: 311.2 to 414.5 hours.)

Log	$\bar{x}$	s	Confidence interval
Common	2.5552	0.0653	2.4930 to 2.6175
Natural	5.8836	0.1504	5.7403 to 6.0270

7.51. The standard deviation for the given data was $s \doteq 0.012224$. With $\alpha = 0.05$, $t = \frac{\bar{x}}{s/\sqrt{15}}$, and df $= 14$, we reject H_0 if $|t| \geq 2.145$, which means $|\bar{x}| \geq (2.145)(s/\sqrt{15})$, or $|\bar{x}| \geq 0.00677$. Assuming $\mu = 0.002$:

$$\begin{aligned} P(|\bar{x}| \geq 0.00677) &= 1 - P(-0.00677 \leq \bar{x} \leq 0.00677) \\ &= 1 - P\left(\frac{-0.00677 - 0.002}{s/\sqrt{15}} \leq \frac{\bar{x} - 0.002}{s/\sqrt{15}} \leq \frac{0.00677 - 0.002}{s/\sqrt{15}}\right) \\ &= 1 - P(-2.78 \leq Z \leq 1.51) \\ &= 1 - (0.9345 - 0.0027) \doteq 0.07 \end{aligned}$$

The power is about 7% against this alternative—not surprising, given the small sample size, and the fact that the difference (0.002) is small relative to the standard deviation.

Note: *Power calculations are often done with software. This may give answers that differ slightly from those found by the method described in the text. Most software does these computations with a "noncentral t distribution" (used in the text for two-sample power problems) rather than a Normal distribution, resulting in more accurate answers. In most situations, the practical conclusions drawn from the power computations are the same regardless of the method used.*

7.52. We will reject H_0 when $t = \frac{\bar{x}}{s/\sqrt{n}} \geq t^*$, where t^* is the appropriate critical value for the chosen sample size. This corresponds to $\bar{x} \geq 15t^*/\sqrt{n}$, so the power against $\mu = 2$ is:

$$\begin{aligned} P(\bar{x} \geq 15t^*/\sqrt{n}) &= P\left(\frac{\bar{x} - 2}{15/\sqrt{n}} \geq \frac{15t^*/\sqrt{n} - 2}{15/\sqrt{n}}\right) \\ &= P\left(Z \geq t^* - \frac{2}{15}\sqrt{n}\right) \end{aligned}$$

For $\alpha = 0.05$, the table on the right shows the power for a variety of sample sizes, and we see that $n \geq 349$ achieves the desired 80% power.

n	t^*	$t^* - \frac{2}{15}\sqrt{n}$	Power
100	1.6604	0.3271	0.3718
200	1.6525	−0.2331	0.5921
300	1.6500	−0.6594	0.7452
340	1.6494	−0.8092	0.7908
345	1.6493	−0.8273	0.7960
346	1.6493	−0.8309	0.7970
347	1.6493	−0.8345	0.7980
348	1.6493	−0.8380	0.7990
349	1.6492	−0.8416	0.8000

7.53. Taking $s = 1.5$ as in Example 7.9, the power for the alternative $\mu = 0.75$ is:

$$P\left(\bar{x} \geq \frac{t^* s}{\sqrt{n}} \text{ when } \mu = 0.75\right) = P\left(\frac{\bar{x} - 0.75}{s/\sqrt{n}} \geq \frac{t^* s/\sqrt{n} - 0.75}{s/\sqrt{n}}\right) = P\left(Z \geq t^* - 0.5\sqrt{n}\right)$$

Using trial-and-error, we find that with $n = 26$, power $\doteq 0.7999$, and with $n = 27$, power $\doteq 0.8139$. Therefore, we need $n > 26$.

7.54. **(a)** Use a two-sided alternative (H_a: $\mu_A \neq \mu_B$) because we (presumably) have no prior suspicion that one design will be better than the other. **(b)** Both sample sizes are the same ($n_1 = n_2 = 15$), so the appropriate degrees of freedom would be df $= 15 - 1 = 14$. **(c)** For a two-sided test at $\alpha = 0.05$, we need $|t| > t^*$, where $t^* = 2.145$ is the 0.025 critical value for a t distribution with df $= 14$.

7.55. Because $2.264 < t < 2.624$ and the alternative is two-sided, Table D tells us that the P-value is $0.02 < P < 0.04$. (Software gives $P = 0.0280$.) That is sufficient to reject H_0 at $\alpha = 0.05$.

7.56. We find $\text{SE}_D \doteq 2.0396$. The options for the 95% confidence interval for $\mu_1 - \mu_2$ are shown on the right. This interval includes fewer values than a 99% confidence interval would (that is, a 99% confidence interval would be wider) because increasing our confidence level means that we need a larger margin of error.

df	t^*	Confidence interval
85.4	1.9881	−14.0550 to −5.9450
49	2.0096	−14.0987 to −5.9013
40	2.021	−14.1220 to -5.8780

7.57. We find $\text{SE}_D \doteq 4.5607$. The options for the 95% confidence interval for $\mu_1 - \mu_2$ are shown on the right. The instructions for this exercise say to use the second approximation (df = 9), in which case we do not reject H_0, because 0 falls in the the 95% confidence interval. Using the first approximation (df = 15.7, typically given by software), the interval is narrower, and we would reject H_0 at $\alpha = 0.05$ against a two-sided alternative. (In fact, $t \doteq -2.193$, for which $0.05 < P < 0.1$ [Table D, df = 9], or $P \doteq 0.0438$ [software, df $\doteq$ 15.7].)

df	t^*	Confidence interval
15.7	2.1236	−19.6851 to −0.3149
9	2.262	−20.3163 to 0.3163

7.58. $$\text{df} = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{\dfrac{(s_1^2/n_1)^2}{n_1 - 1} + \dfrac{(s_2^2/n_2)^2}{n_2 - 1}} = \frac{(12^2/30 + 9^2/25)^2}{\dfrac{(12^2/30)^2}{29} + \dfrac{(9^2/25)^2}{24}} \doteq 52.4738.$$

7.59. SPSS and SAS give both results (the SAS output refers to the unpooled result as the Satterthwaite method), while JMP and Excel show only the unpooled procedures. The pooled t statistic is 1.998, for which $P = 0.0808$.

Note: *When the sample sizes are equal—as in this case—the pooled and unpooled t statistics are equal. (See the next exercise.)*

Both Excel and JMP refer to the unpooled test with the slightly-misleading phrase "assuming unequal variances." The SAS output also implies that the variances are unequal for this method. In fact, unpooled procedures make no *assumptions about the variances.*

Finally, note that both Excel and JMP can do pooled procedures as well as the unpooled procedures that are shown.

7.60. If $n_1 = n_2 = n$, the pooled estimate of the variance is:

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = \frac{(n - 1)s_1^2 + (n - 1)s_2^2}{2(n - 1)} = \frac{s_1^2 + s_2^2}{2}$$

The pooled standard error is therefore $s_p\sqrt{\frac{1}{n} + \frac{1}{n}} = \sqrt{s_p^2 \cdot \frac{2}{n}} = \sqrt{\frac{s_1^2 + s_2^2}{n}} = \sqrt{\frac{s_1^2}{n} + \frac{s_2^2}{n}}$, which is the same as the unpooled standard error.

Note: *The text refers to this as "simple algebra." Bear in mind that some students might consider that phrase to be an oxymoron.*

7.61. (a) Hypotheses should involve μ_1 and μ_2 (population means) rather than $\bar{x}_1$ and $\bar{x}_2$ (sample means). **(b)** The samples are not independent; we would need to compare the 56 males to the 44 females. **(c)** We need P to be small (for example, less than 0.10) to reject H_0. A large P-value like this gives no reason to doubt H_0. **(d)** Assuming the researcher computed the t statistic using $\bar{x}_1 - \bar{x}_2$, a positive value of t does not support H_a. (The one-sided P-value would be 0.982, not 0.018.)

7.62. (a) Because 0 is not in the confidence interval, we would reject H_0 at the 5% level. **(b)** Larger samples generally give smaller margins of error (at the same confidence level, and assuming that the standard deviations for the large and small samples are about the same). One conceptual explanation for this is that larger samples give more information and therefore offer more precise results. A more mathematical explanation: In looking at the formula for a two-sample confidence interval, we see that $\text{SE}_D = \sqrt{s_1^2/n_1 + s_2^2/n_2}$, so that if n_1 and n_2 are increased, the standard error decreases.

Note: *For (a), we can even make some specific statements about t and its P-value: The confidence interval tells us that $\bar{x}_1 - \bar{x}_2 = 1.55$ (halfway between 0.8 and 2.3) and the margin of error is 0.75 (half the width of the interval). As t^* for a 95% confidence interval is at least 1.96, $SE_D = \frac{0.75}{t^*}$ is less than about 0.383 and the t-statistic $t = 1.55/SE_D$ is at least 4.05. (The largest possible value of t, for df = 1, is about 26.3.) A little experimentation with different df reveals that for all df, $P < 0.024$; if df ≥ 10, then $P < 0.001$.*

7.63. (a) We cannot reject H_0: $\mu_1 = \mu_2$ in favor of the two-sided alternative at the 5% level because $0.05 < P < 0.10$ (Table D) or $P \doteq 0.0542$ (software). **(b)** We could reject H_0 in favor of H_a: $\mu_1 < \mu_2$. A negative t-statistic means that $\bar{x}_1 < \bar{x}_2$, which supports the claim that $\mu_1 < \mu_2$, and the one-sided P-value would be half of its value from part (a): $0.025 < P < 0.05$ (Table D) or $P \doteq 0.0271$ (software).

7.64. We find $\text{SE}_D \doteq 3.4792$. The options for the 95% confidence interval for $\mu_1 - \mu_2$ are shown on the right. A 99% confidence interval would include more values (it would be wider) because increasing our confidence level means that we need a larger margin of error.

df	t^*	Confidence interval
87.8	1.9873	−16.9144 to −3.0856
39	2.0227	−17.0374 to −2.9626
30	2.042	−17.1046 to −2.8954

7.65. (a) Stemplots (right) do not look particularly Normal, but they have no extreme outliers or skewness, so t procedures should be reasonably safe. **(b)** The table of summary statistics is below on the left. **(c)** We wish to test H_0: $\mu_N = \mu_S$ versus H_a: $\mu_N < \mu_S$. **(d)** We find $\text{SE}_D \doteq 0.3593$ and $t = -4.303$, so $P \doteq 0.0001$ (df $\doteq 26.5$) or $P < 0.0005$ (df = 13). Either way, we reject H_0. **(e)** The 95% confidence interval for the difference is one of the two options in the table below on the right.

Neutral

0	0000000
0	55
1	000
1	
2	00

Sad

0	0
0	5
1	000
1	555
2	0
2	55
3	00
3	55
4	00

Group	n	$\bar{x}$	s
Neutral	14	\$0.5714	\$0.7300
Sad	17	\$2.1176	\$1.2441

df	t^*	Confidence interval
26.5	2.0538	−2.2842 to −0.8082
13	2.160	−2.3224 to −0.7701

7.66. **(a)** The scores can be examined with either histograms or stemplots. Neither distribution reveals any extreme skewness or outliers, so t procedures should be safe. **(b)** Summary statistics for the two distributions are given below on the left. We find $\text{SE}_D \doteq 0.2891$ and $t = 3.632$, so $P \doteq 0.0008$ (df $\doteq$ 39.5) or $0.001 < P < 0.002$ (df = 19). Either way, we reject H_0. **(c)** The 95% confidence interval for the difference is one of the two options in the table below on the right. **(d)** The hypothesis test and confidence interval suggest that primed individuals had a more positive response to the shampoo label, with an average rating between 0.4 and 1.6 points higher than the unprimed group. (However, priming can only do so much, as the average score in the primed group was only 4 on a scale of 1 to 7.)

Primed		Non-primed	
1		1	00
2	00	2	00
3	000	3	000000000000
4	0000000000	4	000
5	0000000	5	0

Group	n	$\bar{x}$	s
Primed	22	4.00	0.9258
Non-primed	20	2.95	0.9445

df	t^*	Confidence interval
39.5	2.0220	0.4655 to 1.6345
19	2.093	0.4450 to 1.6550

7.67. **(a)** The female means and standard deviations are $\bar{x}_F \doteq 4.0791$ and $s_F \doteq 0.9861$; for males, they are $\bar{x}_M \doteq 3.8326$ and $s_M \doteq 1.0677$. **(b)** Both distributions are somewhat skewed to the left. This can be seen by constructing a histogram, but is also evident in the data table in the text by noting the large numbers of "4" and "5" ratings for both genders. However, because the ratings range from 1 to 5, there are no outliers, so the t procedures should be safe. **(c)** We find $\text{SE}_D \doteq 0.0851$ and $t \doteq 2.898$, for which $P \doteq 0.0040$ (df $\doteq$ 402.2) or $0.002 < P < 0.005$ (df = 220). Either way, there is strong evidence of a difference in satisfaction. **(d)** The 95% confidence interval for the difference is one of the three options in the table on the right—roughly 0.08 to 0.41. **(e)** While we have evidence of a difference in mean ratings, it might not be as large as 0.25.

df	t^*	Confidence interval
402.2	1.9659	0.0793 to 0.4137
220	1.9708	0.0788 to 0.4141
100	1.984	0.0777 to 0.4153

7.68. **(a)** For testing H_0: $\mu_{\text{LC}} = \mu_{\text{LF}}$ versus H_a: $\mu_{\text{LC}} \neq \mu_{\text{LF}}$, we have $\text{SE}_D \doteq 6.7230$ and $t \doteq 4.165$, so $P < 0.0001$ (df $\doteq$ 62.1) or $P < 0.001$ (df = 31). Either way, we clearly reject H_0. **(b)** It might be that the moods of subjects who dropped out differed from the moods of those who stayed; in particular, it seems reasonable to suspect that those who dropped out had higher TMDS scores.

7.69. **(a)** Assuming we have SRSs from each population, use of two-sample t procedures seems reasonable. (We cannot assess Normality, but the large sample sizes would overcome most problems.)

df	t^*	Confidence interval
76.1	1.9916	−11.7508 to 15.5308
36	2.0281	−12.0005 to 15.7805
30	2.042	−12.0957 to 15.8757

(b) We wish to test H_0: $\mu_f = \mu_m$ versus H_a: $\mu_f \neq \mu_m$. **(c)** We find $\text{SE}_D \doteq 6.8490$ mg/dl. The test statistic is $t \doteq 0.276$, with df $\doteq 76.1$ (or 36—use 30 for Table D), for which $P \doteq 0.78$. We have no reason to believe that male and female cholesterol levels are different. **(d)** The options for the 95% confidence interval for $\mu_f - \mu_m$ are shown on the right. **(e)** It might not be appropriate to treat these students as SRSs from larger populations.

Note: *Because t distributions are more spread out than Normal distributions, a t-value that would not be significant for a Normal distribution (such as 0.276) cannot possibly be significant when compared to a t distribution.*

7.70. Considering the LDL cholesterol levels, we test H_0: $\mu_f = \mu_m$ versus H_a: $\mu_f < \mu_m$. We find $\text{SE}_D \doteq 6.2087$ mg/dl, so the test statistic is $t \doteq -2.104$, with df $\doteq 70.5$ (or 36—use 30 for Table D), for which $P \doteq 0.0195$. We have enough evidence to conclude LDL cholesterol levels are higher for males than for females.

Note: *This t statistic was computed by subtracting the male mean from the female mean. Reversing the order of subtraction would make t positive, but would not change the P-value or the conclusion.*

7.71. **(a)** The distribution cannot be Normal because all numbers are integers. **(b)** The t procedures should be appropriate because we have two large samples with no outliers. **(c)** We will test H_0: $\mu_I = \mu_C$ versus H_a: $\mu_I > \mu_C$ (or $\mu_I \neq \mu_C$). The one-sided alternative reflects the researchers' (presumed) belief that the intervention would increase scores on the test. The two-sided alternative allows for the possibility that the intervention might have had a negative effect.

df	t^*	Confidence interval
354.0	1.9667	0.5143 to 0.9857
164	1.9745	0.5134 to 0.9866
100	1.984	0.5122 to 0.9878

(d) $\text{SE}_D = \sqrt{s_I^2/n_I + s_C^2/n_C} \doteq 0.1198$ and $t = (\bar{x}_I - \bar{x}_C)/\text{SE}_D \doteq 6.258$. Regardless of how we compute degrees of freedom (df $\doteq 354$ or 164), the P-value is very small: $P < 0.0001$. We reject H_0 and conclude that the intervention increased test scores. **(e)** The interval is $\bar{x}_I - \bar{x}_C \pm t^*\text{SE}_D$; the value of t^* depends on the df (see the table), but note that in every case the interval rounds to 0.51 to 0.99. **(f)** The results for this sample may not generalize well to other areas of the country.

7.72. **(a)** The distribution cannot be Normal, because all numbers are integers. **(b)** The t procedures should be appropriate because we have two large samples with no outliers. **(c)** Again, we test H_0: $\mu_I = \mu_C$ versus H_a: $\mu_I > \mu_C$ (or $\mu_I \neq \mu_C$). The one-sided alternative reflects the researchers' (presumed) belief that the intervention would increase self-efficacy scores. The two-sided alternative allows for the possibility that the intervention might have had a negative effect. **(d)** $\text{SE}_D = \sqrt{s_I^2/n_I + s_C^2/n_C} \doteq 0.1204$ and $t = (\bar{x}_I - \bar{x}_C)/\text{SE}_D \doteq 3.571$. Regardless of how we compute degrees of freedom (df $\doteq$ 341.8 or 164), the (one-sided) P-value is about 0.0002. We reject H_0 and conclude that the intervention increased self-efficacy scores. **(e)** The interval is $\bar{x}_I - \bar{x}_C \pm t^*\text{SE}_D$; the value of t^* depends on the df (see the table), but in every case the interval rounds to 0.19 to 0.67. **(f)** As in the previous exercise, the results for this sample may not generalize well to other areas of the country.

df	t^*	Confidence interval
341.8	1.9669	0.1932 to 0.6668
164	1.9745	0.1922 to 0.6678
100	1.984	0.1911 to 0.6689

7.73. **(a)** This may be near enough to an SRS, if this company's working conditions were similar to that of other workers. **(b)** $\text{SE}_D \doteq 0.7626$; regardless of how we choose df, the interval rounds to 9.99 to 13.01 mg.y/m^3. **(c)** A one-sided alternative would seem to be reasonable here; specifically, we would likely expect that the mean exposure for outdoor workers would be lower. For testing H_0, we find $t = 15.08$, for which $P < 0.0001$ with either df $= 137$ or 114 (and for either a one- or a two-sided alternative). We have strong evidence that outdoor concrete workers have lower dust exposure than the indoor workers. **(d)** The sample sizes are large enough that skewness should not matter.

df	t^*	Confidence interval
137.1	1.9774	9.9920 to 13.0080
114	1.9810	9.9893 to 13.0107
100	1.984	9.9870 to 13.0130

7.74. With the given standard deviations, $\text{SE}_D \doteq 0.2653$; regardless of how we choose df, a 95% confidence interval for the difference in means rounds to 4.37 to 5.43 mg.y/m^3. With the null hypothesis H_0: $\mu_i = \mu_o$ (and either a one- or two-sided alternative, as in the previous exercise), we find $t = 18.47$, for which $P < 0.0001$ regardless of df and the chosen alternative. We have strong evidence that outdoor concrete workers have lower respirable dust exposure than the indoor workers.

df	t^*	Confidence interval
121.5	1.9797	4.3747 to 5.4253
114	1.9810	4.3744 to 5.4256
100	1.984	4.3736 to 5.4264

7.75. To find a confidence interval $(\bar{x}_1 - \bar{x}_2) \pm t^*\text{SE}_D$, we need one of the following:

- Sample sizes and standard deviations—in which case we could find the interval in the usual way
- t and df—because $t = (\bar{x}_1 - \bar{x}_2)/\text{SE}_D$, so we could compute $\text{SE}_D = (\bar{x}_1 - \bar{x}_2)/t$ and use df to find t^*
- df and a more accurate P-value—from which we could determine t, and then proceed as above

The confidence interval could give us useful information about the magnitude of the difference (although with such a small P-value, we do know that a 95% confidence interval would not include 0).

df	t^*	Confidence interval
7.8	2.3159	−16.4404 to 3.8404
4	2.776	−18.4551 to 5.8551

7.76. **(a)** The 68–95–99.7 rule suggests that the distributions are not Normal: If they were Normal, then (for example) 95% of 7-to-10-year-olds drink between −13.2 and 29.6 oz of sweetened drinks per day. As negative numbers do not make sense (unless some children are regurgitating sweetened drinks), the distributions must be right-skewed. **(b)** We find $\text{SE}_D \doteq 4.3786$ and $t \doteq -1.439$, with either df $\doteq 7.8$ ($P = 0.1890$) or df $= 4$ ($P = 0.2236$). We do not have enough evidence to reject H_0. **(c)** The possible 95% confidence intervals are given in the table. **(d)** Because the distributions are not Normal and the samples are small, the t procedures are questionable for these data. **(e)** Because this group is not an SRS—and indeed might not be random in any way—we would have to be very cautious about extending these results to other children.

7.77. This is a matched pairs design; for example, Monday hits are (at least potentially) not independent of one another. The correct approach would be to use one-sample t methods on the seven differences (Monday hits for design 1 minus Monday hits for design 2, Tuesday/1 minus Tuesday/2, and so on).

df	t^*	Confidence interval
12.7	2.1651	0.6667 to 2.9333
9	2.262	0.6160 to 2.9840

7.78. **(a)** Results for this randomization will depend on the technique used. **(b)** $\text{SE}_D \doteq 0.5235$, and the options for the 95% confidence interval are given on the right. **(c)** Because 0 falls outside the 95% confidence interval, the P-value is less than 0.05, so we would reject H_0. (For reference, $t \doteq 3.439$ and the actual P-value is either 0.0045 or 0.0074, depending on which df we use.)

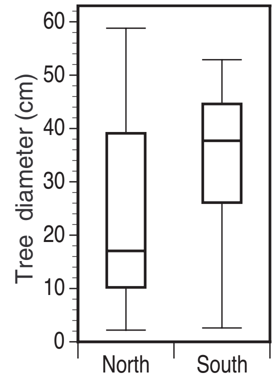
7.79. The next 10 employees who need screens might not be an independent group—perhaps they all come from the same department, for example. Randomization reduces the chance that we end up with such unwanted groupings.

df	t^*	Confidence interval
121.5	1.9797	1.7865 to 2.8935
60	2.000	1.7808 to 2.8992

7.80. **(a)** The null hypothesis is $\mu_1 = \mu_2$; the alternative can be either two- or one-sided. (It might be a reasonable expectation that $\mu_1 > \mu_2$.) We find $\text{SE}_D \doteq 0.2796$ and $t = 8.369$. Regardless of df and H_a, the conclusion is the same: P is very small, and we conclude that *WSJ* ads are more trustworthy. **(b)** Possible 95% confidence intervals are given in the table; both place the difference in trustworthiness at between about 1.8 and 2.9 points. **(c)** Advertising in *WSJ* is seen as more reliable than advertising in the *National Enquirer*—a conclusion that probably comes as a surprise to no one.

7.81. **(a)** Stemplots, boxplots, and five-number summaries (in cm) are shown on the right. The north distribution is right-skewed, while the south distribution is left-skewed. **(b)** The methods of this section seem to be appropriate in spite of the skewness because the sample sizes are relatively large, and there are no outliers in either distribution. **(c)** We test H_0: $\mu_n = \mu_s$ versus H_a: $\mu_n \neq \mu_s$; we should use a two-sided alternative because we have no reason (before looking at the data) to expect a difference in a particular direction. **(d)** The means and standard deviations are $\bar{x}_n = 23.7$, $s_n \doteq 17.5001$, $\bar{x}_s = 34.5\overline{3}$, and $s_s \doteq 14.2583$ cm. Then $\text{SE}_D \doteq 4.1213$, so $t = -2.629$ with df $= 55.7$ ($P = 0.011$) or df $= 29$ ($P = 0.014$). We conclude that the means are different (specifically, the south mean is greater than the north mean). **(e)** See the table for possible 95% confidence intervals.

North		South
43322	0	2
65	0	57
443310	1	2
955	1	8
	2	13
8755	2	689
0	3	2
996	3	566789
43	4	003444
6	4	578
4	5	0112
85	5	

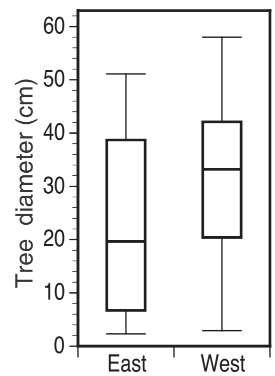


	Min	Q_1	M	Q_3	Max
North	2.2	10.2	17.05	39.1	58.8
South	2.6	26.1	37.70	44.6	52.9

df	t^*	Confidence interval
55.7	2.0035	−19.0902 to −2.5765
29	2.045	−19.2614 to −2.4053

7.82. **(a)** Stemplots, boxplots, and five-number summaries (in cm) are shown on the right. The east distribution is right-skewed, while the west distribution is left-skewed. **(b)** The methods of this section seem to be appropriate in spite of the skewness because the sample sizes are relatively large, and there are no outliers in either distribution. **(c)** We test H_0: $\mu_e = \mu_w$ versus H_a: $\mu_e \neq \mu_w$; we should use a two-sided alternative because we have no reason (before looking at the data) to expect a difference in a particular direction. **(d)** The means and standard deviations are $\bar{x}_e = 21.71\overline{6}$, $s_e \doteq 16.0743$, $\bar{x}_w = 30.28\overline{3}$, and $s_w \doteq 15.3314$ cm. Then $\text{SE}_D \doteq 4.0556$, so $t = -2.112$ with df $= 57.8$ ($P = 0.0390$) or df $= 29$ ($P = 0.0434$). We conclude that the means are different at $\alpha = 0.05$ (specifically, the west mean is greater than the east mean). **(e)** See the table for possible 95% confidence intervals.

East		West
222	0	233
9566655	0	
3100	1	11
7	1	78
33222	2	0011
	2	55
11	3	00
98	3	555669
333	4	023444
86	4	1
1	5	78



	Min	Q_1	M	Q_3	Max
East	2.3	6.7	19.65	38.7	51.1
West	2.9	20.4	33.20	42.1	58.0

df	t^*	Confidence interval
57.8	2.0018	−16.6852 to −0.4481
29	2.045	−16.8604 to −0.2730

7.83. **(a)** $SE_D \doteq 1.9686$. Answers will vary with the df used (see the table), but the interval is roughly -1 to 7 units. **(b)** Because of random fluctuations between stores, we might (just by chance) have seen a rise in the average number of units sold even if actual mean sales had remained unchanged. (Based on the confidence interval, mean sales might have even dropped slightly.)

df	t^*	Confidence interval
122.5	1.9795	−0.8968 to 6.8968
54	2.0049	−0.9468 to 6.9468
50	2.009	−0.9549 to 6.9549

7.84. **(a)** Good statistical practice dictates that the alternative hypothesis should be chosen without looking at the data; we should only choose a one-sided alternative if we have some reason to expect it *before* looking at the sample results. **(b)** The correct P-value is twice that reported for the one-tailed test: $P = 0.12$.

7.85. **(a)** We test H_0: $\mu_b = \mu_f$; H_a: $\mu_b > \mu_f$. $SE_D \doteq 0.5442$ and $t = 1.654$, for which $P = 0.0532$ (df $= 37.6$) or 0.0577 (df $= 18$); there is not quite enough evidence to reject H_0 at $\alpha = 0.05$. **(b)** The confidence interval depends on the degrees of freedom used; see the table. **(c)** We need two independent SRSs from Normal populations.

df	t^*	Confidence interval
37.6	2.0251	−0.2021 to 2.0021
18	2.101	−0.2434 to 2.0434

7.86. See the solution to Exercise 7.65 for a table of means and standard deviations. The pooled standard deviation is $s_p \doteq 1.0454$, so the pooled standard error is $s_p\sqrt{1/14 + 1/17} \doteq 0.3773$. The test statistic is $t \doteq -4.098$ with df $= 29$, for which $P = 0.0002$, and the 95% confidence interval (with $t^* \doteq 2.045$) is -2.3178 to -0.7747.

In the solution to Exercise 7.65, we reached the same conclusion on the significance test ($t \doteq -4.303$ and $P \doteq 0.0001$) and the confidence interval was quite similar (roughly -2.3 to -0.8).

7.87. See the solution to Exercise 7.66 for a table of means and standard deviations. The pooled standard deviation is $s_p \doteq 0.9347$, so the pooled standard error is $s_p\sqrt{1/22 + 1/20} \doteq 0.2888$. The test statistic is $t \doteq 3.636$ with df $= 40$, for which $P = 0.0008$, and the 95% confidence interval (with $t^* \doteq 2.021$) is 0.4663 to 1.6337.

In the solution to Exercise 7.66, we reached the same conclusion on the significance test ($t \doteq 3.632$ and $P \doteq 0.0008$) and the confidence interval (using the more-accurate df $\doteq 39.5$) was quite similar: 0.4655 to 1.6345.

7.88. See the solution to Exercise 7.67 for means and standard deviations. The pooled standard deviation is $s_p \doteq 1.0129$, so the pooled standard error is $s_p\sqrt{1/468 + 1/221} \doteq 0.0827$. The test statistic is $t \doteq 2.981$ with df $= 687$, for which $P = 0.0030$ (or, using Table D, $0.002 < P < 0.005$). The 95% confidence interval is one of the two entries in the table on the right.

df	t^*	Confidence interval
687	1.9634	0.0842 to 0.4088
100	1.984	0.0825 to 0.4105

In the solution to Exercise 7.67, we reached the same conclusion on the significance test ($t \doteq 2.898$ and $P \doteq 0.0040$). The confidence intervals were slightly wider, but similar; using the more-accurate df $\doteq 402.2$, the interval was 0.0793 to 0.4137. (The other intervals were wider than this.)

7.89. See the solution to Exercise 7.81 for means and standard deviations. The pooled standard deviation is $s_p \doteq 15.9617$, and the standard error is $\text{SE}_D \doteq 4.1213$. For the significance test, $t = -2.629$, df $= 58$, and $P = 0.0110$, so we have fairly strong evidence (though not quite significant at $\alpha = 0.01$) that the south mean is greater than the north mean. Possible answers for the confidence interval (with software, and with Table D) are given in the table. All results are similar to those found in Exercise 7.81.

df	t^*	Confidence interval
58	2.0017	-19.0830 to -2.5837
50	2.009	-19.1130 to -2.5536

Note: *If $n_1 = n_2$ (as in this case), the standard error and t statistic are the same for the usual and pooled procedures. The degrees of freedom will usually be different (specifically, df is larger for the pooled procedure, unless $s_1 = s_2$ and $n_1 = n_2$).*

7.90. Testing the same hypotheses as in that example (H_0: $\mu_1 = \mu_2$ versus H_a: $\mu_1 \neq \mu_2$), we have pooled standard deviation $s_p \doteq 8.1772$ so that $\text{SE}_D \doteq 1.2141$ and $t \doteq 4.777$. With either df $= 236$ or 100, we find that $P < 0.001$, so we have very strong evidence that mean systolic blood pressure is higher in low sleep efficiency children. This is nearly identical to the results of the unpooled analysis ($t = 4.18$, $P < 0.001$).

7.91. With $s_n \doteq 17.5001$, $s_s \doteq 14.2583$, and $n_n = n_s = 30$, we have $s_n^2/n_n \doteq 10.2085$ and $s_s^2/n_s \doteq 6.7767$, so:

$$\text{df} = \frac{\left(\frac{s_n^2}{n_n} + \frac{s_s^2}{n_s}\right)^2}{\frac{1}{n_n - 1}\left(\frac{s_n^2}{n_n}\right)^2 + \frac{1}{n_s - 1}\left(\frac{s_s^2}{n_s}\right)^2} \doteq \frac{(10.2085 + 6.7767)^2}{\frac{1}{29}(10.2085^2 + 6.7767^2)} \doteq 55.7251$$

7.92. With $s_e \doteq 16.0743$, $s_w \doteq 15.3314$, and $n_e = n_w = 30$, we have $s_e^2/n_e \doteq 8.6128$ and $s_w^2/n_w \doteq 7.8351$, so:

$$\text{df} = \frac{\left(\frac{s_e^2}{n_e} + \frac{s_w^2}{n_w}\right)^2}{\frac{1}{n_e - 1}\left(\frac{s_e^2}{n_e}\right)^2 + \frac{1}{n_w - 1}\left(\frac{s_w^2}{n_w}\right)^2} \doteq \frac{(8.6128 + 7.8351)^2}{\frac{1}{29}(8.6128^2 + 7.8351^2)} \doteq 57.8706$$

7.93. **(a)** With $s_i \doteq 7.8$, $n_i = 115$, $s_o \doteq 3.4$, and $n_o = 220$, we have $s_i^2/n_i \doteq 0.5290$ and $s_o^2/n_o \doteq 0.05455$, so:

$$\text{df} \doteq \frac{(0.5290 + 0.05455)^2}{\frac{0.5290^2}{114} + \frac{0.05455^2}{219}} \doteq 137.0661$$

	df	t^*	Confidence interval
Part (d)	333	1.9671	10.2931 to 12.7069
	100	1.984	10.2827 to 12.7173
Part (e)	333	1.9671	4.5075 to 5.2925
	100	1.984	4.5042 to 5.2958

(b) $s_p = \sqrt{\frac{(n_i-1)s_i^2+(n_o-1)s_o^2}{n_i+n_o-2}} \doteq 5.3320$, which is slightly closer to s_o (the standard deviation from the larger sample). **(c)** With no assumption of equality, $\text{SE}_1 = \sqrt{s_i^2/n_i + s_o^2/n_o} \doteq 0.7626$. With the pooled method, $\text{SE}_2 = s_p\sqrt{1/n_i + 1/n_o} \doteq 0.6136$. **(d)** With the pooled standard deviation, $t \doteq 18.74$ and df $= 333$, for which $P < 0.0001$, and the 95% confidence interval is as shown in the table. With the smaller standard error, the t value is larger (it had been 15.08), and the confidence interval is narrower. The P-value is also smaller (although both are less than 0.0001). **(e)** With $s_i \doteq 2.8$, $n_i = 115$, $s_o \doteq 0.7$, and $n_o = 220$, we have $s_i^2/n_i \doteq 0.06817$ and $s_o^2/n_o \doteq 0.002227$, so:

$$\text{df} \doteq \frac{(0.06817 + 0.002227)^2}{\frac{0.06817^2}{114} + \frac{0.002227^2}{219}} \doteq 121.5030$$

The pooled standard deviation is $s_p \doteq 1.7338$; the standard errors are $\text{SE}_1 = 0.2653$ (with no assumptions) and $\text{SE}_2 = 0.1995$ (assuming equal standard deviations). The pooled t is 24.56 (df $= 333$, $P < 0.0001$), and the 95% confidence intervals are shown in the table. The pooled and usual t procedures compare similarly to the results for part (d): With the pooled procedure, t is larger, and the interval is narrower.

7.94. We have $n_1 = n_2 = 5$. **(a)** For a two-sided test with df $= 4$, the critical value is $t^* = 2.776$. **(b)** With the pooled procedures, df $= 8$ and the critical value is $t^* = 2.306$. **(c)** The smaller critical value with the pooled approach means that a smaller t-value (that is, weaker evidence) is needed to reject H_0.

Note: *When software is available, we use the more accurate degrees of freedom for the standard approach. In this case, pooling typically is less beneficial; for this example, the software output shown in Figure 7.14 shows that df $\doteq 7.98$ for the unpooled approach.*

7.95. **(a)** From an $F(15, 22)$ distribution with $\alpha = 0.05$, $F^* = 2.15$. **(b)** Because $F = 2.45$ is greater than the 5% critical value, but less than the 2.5% critical value ($F^* = 2.50$), we know that P is between $2(0.025) = 0.05$ and $2(0.05) = 0.10$. (Software tells us that $P = 0.055$.) $F = 2.45$ is significant at the 10% level but not at the 5% level.

7.96. The power would be higher. Larger differences are easier to detect; that is, when $\mu_1 - \mu_2$ is more than 5, there is a greater chance that the test statistic will be significant.

Note: *In fact, as the table on the right shows, if we repeat the computations of Example 7.23 with larger values of $\mu_1 - \mu_2$, the power increases rapidly.*

$\mu_1 - \mu_2$	Power
5	0.7965
6	0.9279
7	0.9817
8	0.9968
9	0.9996

7.97. The power would be smaller. A larger value of σ means that large differences between the sample means would arise more often by chance so that, if we observe such a difference, it gives less evidence of a difference in the population means.

Note: *The table on the right shows the decrease in the power as σ increases.*

σ	Power
7.4	0.7965
7.5	0.7844
7.6	0.7722
7.7	0.7601
7.8	0.7477

7.98. (a) $F = \frac{9.1}{3.5} = 2.6$. **(b)** For 17 and 8 degrees of freedom (using 15 and 8 in Table E), we need $F > 4.10$ to reject H_0 at the 5% level with a two-sided alternative. **(c)** We cannot conclude that the standard deviations are different. (Software gives $P = 0.1716$.)

7.99. The test statistic is $F = \left(\frac{9.9}{7.5}\right)^2 \doteq 1.742$, with df 60 and 176. The two-sided P-value is 0.0057, so we can reject H_0 and conclude that the standard deviations are different. We do not know if the distributions are Normal, so this test may not be reliable.

7.100. The test statistic is $F = \left(\frac{13.75}{7.94}\right)^2 \doteq 2.9989$, with df 70 and 36. The two-sided P-value is 0.0005, so we have strong evidence that the standard deviations are different. However, this test assumes that the underlying distributions are Normal; if this is not true, then the conclusion may not be reliable.

7.101. The test statistic is $F = \frac{1.16^2}{1.15^2} \doteq 1.0175$ with df 211 and 164. Table E tells us that $P > 0.20$, while software gives $P = 0.9114$. The distributions are not Normal ("total score was an integer between 0 and 6"), so the test may not be reliable (although with s_1 and s_2 so close, the conclusion is probably correct). To reject at the 5% level, we would need $F > F^*$, where $F^* = 1.46$ (using df 120 and 100 from Table E) or $F^* = 1.3392$ (using software). As $F = s_2^2/s_1^2$, we would need $s_2^2 > s_1^2 F^*$, or $s_2 > 1.15\sqrt{F^*}$, which is about 1.3896 (Table E) or 1.3308 (software).

7.102. The test statistic is $F = \frac{1.19^2}{1.12^2} \doteq 1.1289$ with df 164 and 211. Table E tells us that $P > 0.2$, while software gives $P = 0.4063$. We cannot conclude that the standard deviations are different. The distributions are not Normal (because all responses are integers from 1 to 5), so the test may not be reliable.

7.103. The test statistic is $F = \frac{7.8^2}{3.7^2} \doteq 5.2630$ with df 114 and 219. Table E tells us that $P < 0.002$, while software gives $P < 0.0001$; we have strong evidence that the standard deviations differ. The authors described the distributions as somewhat skewed, so the Normality assumption may be violated.

7.104. The test statistic is $F = \frac{2.8^2}{0.7^2} = 16$ with df 114 and 219. Table E tells us that $P < 0.002$, while software gives $P < 0.0001$; we have strong evidence that the standard deviations differ. We have no information about the Normality of the distributions, so it is difficult to determine how reliable these conclusions are. (We can observe that for Exercise 7.73, $\bar{x}_1 - 3s_1$ and $\bar{x}_2 - 3s_2$ were both negative, hinting at the skewness of those distributions. For Exercise 7.74, this is not the case, suggesting that these distributions might not be as skewed.)

7.105. The test statistic is $F \doteq \frac{17.5001^2}{14.2583^2} \doteq 1.5064$ with df 29 and 29. Table E tells us that $P > 0.2$, while software gives $P = 0.2757$; we cannot conclude that the standard deviations differ. The stemplots and boxplots of the north/south distributions in Exercise 7.81 do not appear to be Normal (both distributions were skewed), so the results may not be reliable.

7.106. The test statistic is $F \doteq \frac{16.0743^2}{15.3314^2} \doteq 1.0993$ with df 29 and 29. Table E tells us that $P > 0.2$, while software gives $P = 0.8006$; we cannot conclude that the standard deviations differ. The stemplots and boxplots of the east/west distributions in Exercise 7.82 do not appear to be Normal (both distributions were skewed), so the results may not be reliable.

7.107. (a) To test H_0: $\sigma_1 = \sigma_2$ versus H_a: $\sigma_1 \neq \sigma_2$, we find $F \doteq \frac{7.1554^2}{6.7676^2} \doteq 1.1179$. We do not reject H_0. **(b)** With an $F(4, 4)$ distribution with a two-sided alternative, we need the critical value for $p = 0.025$: $F^* = 9.60$. The table on the right gives the critical values for other sample sizes. With such small samples, this is a very low-power test; large differences between σ_1 and σ_2 would rarely be detected.

n	F^*
5	9.60
4	15.44
3	39.00
2	647.79

7.108. (a) For two samples of size 20, we have noncentrality parameter

$$\delta = \frac{10}{20\sqrt{2/20}} \doteq 1.5811$$

The power is about 0.33 (using the Normal approximation) or 0.34 (software); see the table on the right. **(b)** With $n_1 = n_2 = 60$, we have $\delta \doteq 2.7386$, df $= 118$, and $t^* = 1.9803$ (or 1.984 for df $= 100$ from Table D). The approximate power is about 0.78 (details in the table on the right). **(c)** Samples of size 60 would give a reasonably good chance of detecting a difference of 20 cm.

n	δ	t^*	Power (Normal)	Power (software)
20	1.5811	2.0244	0.3288	0.3379
60	2.7386	1.9803	0.7759	0.7753

7.109. The four standard deviations from Exercises 7.81 and 7.82 are $s_n \doteq 17.5001$, $s_s \doteq 14.2583$, $s_e \doteq 16.0743$, and $s_w \doteq 15.3314$ cm. Using a larger σ for planning the study is advisable because it provides a conservative (safe) estimate of the power. For example, if we choose a sample size to provide 80% power and the true σ is smaller than that used for planning, the actual power of the test is greater than the desired 80%.

Results of additional power computations depend on what students consider to be "other reasonable values of σ." Shown in the table are some possible answers using the Normal approximation. (Powers computed using the noncentral t distribution are slightly greater.)

	Power with $n =$	
σ	20	60
15	0.5334	0.9527
16	0.4809	0.9255
17	0.4348	0.8928
18	0.3945	0.8560

7.110. **(a)** The noncentrality parameter is $\delta = \frac{1.5}{1.6\sqrt{2/65}} \doteq 5.3446$. With such a large value of δ, the value of t^* (1.9787 for df = 128, or 1.984 for df = 100 from Table D) does not matter very much. The Normal approximation for the power is $P(Z > t^* - \delta) \doteq 0.9996$ for either choice of t^*. Software gives the same result. **(b)** For samples of size 100, $\delta \doteq 6.6291$, and once again the value of t^* makes little difference; the power is very close to 1 (using the Normal approximation or software). **(c)** Because the effect is large relative to the standard deviation, small samples are sufficient. (Even samples of size 20 will detect this difference with probability 0.8236.)

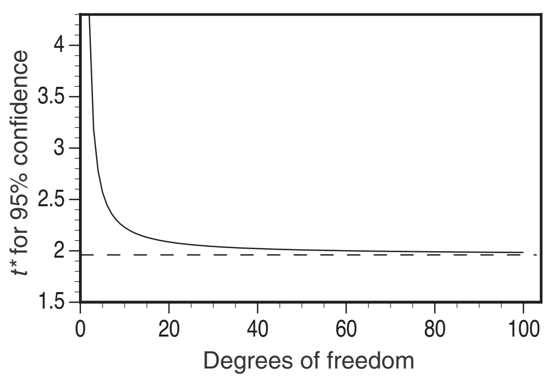
7.111. The mean is $\bar{x} = 140.5$, the standard deviation is $s \doteq 13.58$, and the standard error of the mean is $s_{\bar{x}} \doteq 6.79$. It would not be appropriate to construct a confidence interval because we cannot consider these four scores to be an SRS.

7.112. To support the alternative $\mu_1 > \mu_2$, we need to see $\bar{x}_1 > \bar{x}_2$, so that $t = (\bar{x}_1 - \bar{x}_2)/\text{SE}_D$ must be positive. **(a)** If $t = 2.08$, the one-sided P-value is half of the reported two-sided value ($P = 0.04$), so we reject H_0 at $\alpha = 0.05$. **(b)** $t = -2.08$ does not support H_a; the one-sided P-value is 0.96. We do not reject H_0 at $\alpha = 0.05$ (or any reasonable choice of α).

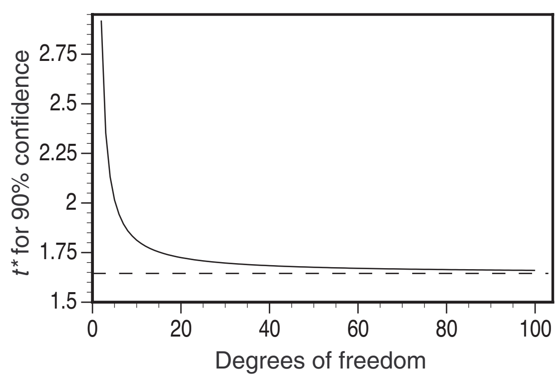
7.113. The plot (below, left) shows that t^* approaches 1.96 as df increases.

7.114. The plot (below, right) shows that t^* approaches 1.645 as df increases.

For 7.113



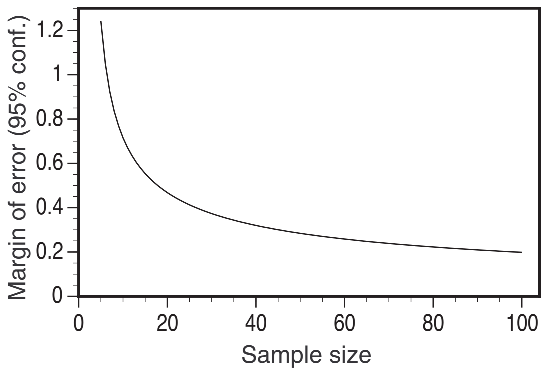
For 7.114



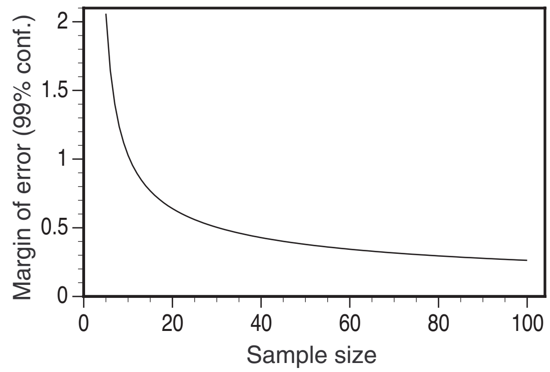
7.115. The margin of error is $t^*/\sqrt{n}$, using t^* for df $= n - 1$ and 95% confidence. For example, when $n = 5$, the margin of error is 1.2417, and when $n = 10$, it is 0.7154, and for $n = 100$, it is 0.1984. As we see in the plot (next page, left), as sample size increases, margin of error decreases (toward 0, although it gets there very slowly).

7.116. The margin of error is $t^*/\sqrt{n}$, using t^* for df $= n - 1$ and 99% confidence. For example, when $n = 5$, the margin of error is 2.0590, and when $n = 10$, it is 1.0277, and for $n = 100$, it is 0.2626. As we see in the plot (next page, right), as sample size increases, margin of error decreases (toward 0, although it gets there very slowly).

For 7.115



For 7.116



7.117. **(a)** Use two independent samples (students that live in the dorms, and those that live elsewhere). **(b)** Use a matched pairs design: Take a sample of college students, and have each subject rate the appeal of each label design. **(c)** Take a single sample of college students, and ask them to rate the appeal of the product.

7.118. **(a)** Take a single sample of customers, and record the age of each subject. **(b)** Use two independent samples (this year's sample, and last year's sample). **(c)** Use a matched pairs design: Ask each customer in the sample to rate each floor plan.

7.119. **(a)** To test H_0: $\mu = 1.5$ versus H_a: $\mu < 1.5$, we have $t = \frac{1.20-1.5}{1.81/\sqrt{200}} \doteq -2.344$ with df $= 199$, for which $P \doteq 0.0100$. We can reject H_0 at the 5% significance level. **(b)** From Table D, use df $= 100$ and $t^* = 1.984$, so the 95% confidence interval for μ is

$$1.20 \pm 1.984\left(\frac{1.81}{\sqrt{200}}\right) = 1.20 \pm 0.2539 = 0.9461 \text{ to } 1.4539 \text{ violations}$$

(With software, the interval is 0.9476 to 1.4524.) **(c)** While the significance test lets us conclude that there were fewer than 1.5 violations (on the average), the confidence interval gives us a range of values for the mean number of violations. **(d)** We have a large sample ($n = 200$), and the limited range means that there are no extreme outliers, so t procedures should be safe.

7.120. **(a)** The prior studies give us reason to expect greater improvement from those who used the computer, so we test H_0: $\mu_C = \mu_D$ versus H_a: $\mu_C > \mu_D$. **(b)** With $\text{SE}_D \doteq 0.7527$, the test statistic is $t \doteq 2.79$, so $P \doteq 0.0027$ (df $\doteq 484.98$) or $0.0025 < P < 0.005$ (df $= 241$; use df $= 100$ in Table D). Either way, we reject H_0 at the $\alpha = 0.05$ level. **(c)** While we have strong evidence that the *mean* improvement is greater using computerized training, we cannot draw this conclusion about an *individual's* response.

7.121. **(a)** The mean difference in body weight change (with wine minus without wine) was $\bar{x}_1 = 0.4 - 1.1 = -0.7$ kg, with standard error $\text{SE}_1 = 8.6/\sqrt{14} \doteq 2.2984$ kg. The mean difference in caloric intake was $\bar{x}_2 = 2589 - 2575 = 14$ cal, with $\text{SE}_2 = 210/\sqrt{14} \doteq 56.1249$ cal. **(b)** The t statistics $t_i = \bar{x}_i/\text{SE}_i$, both with df $= 13$, are $t_1 = -0.3046$ ($P_1 = 0.7655$) and $t_2 = 0.2494$ ($P_2 = 0.8069$). **(c)** For df $= 13$, $t^* = 2.160$, so the 95% confidence intervals $\bar{x}_i \pm t^*\text{SE}_i$ are -5.6646 to 4.2646 kg (-5.6655 to 4.2655 with software) and -107.2297 to 135.2297 cal (-107.2504 to 135.2504 with software).

(d) Students might note a number of factors in their discussions; for example, all subjects were males, weighing 68 to 91 kg (about 150 to 200 lb), which may limit how widely we can extend these conclusions.

7.122. For both entrée and wine, we test H_0: $\mu_C = \mu_N$ versus H_a: $\mu_C > \mu_N$, because the question asked suggests an expectation that consumption would be greater when the wine was identified as coming from California.

df	t^*	Confidence interval
29.3	2.0442	1.2975 to 120.3025
14	2.145	−1.6364 to 123.2364

For the entrée, $SE_D \doteq 29.1079$ and $t \doteq 2.089$, for which $P \doteq 0.0227$ (df $\doteq$ 29.3) or $0.025 < P < 0.05$ (df = 14). For wine consumption, we will not reject H_0, because the sample means are not consistent with H_a: $SE_D \doteq 5.2934$, $t \doteq -1.814$, and $P > 0.5$.

Having rejected H_0 for the entrée, a good second step is to give a confidence interval for the size of the difference; the two options for this interval are in the table on the right. Note that with the conservative degrees of freedom (df = 14), the interval includes 0, because the one-sided P-value was greater than 0.025.

7.123. How much a person eats or drinks may depend on how many people he or she is sitting with. This means that the individual customers within each wine-label group probably cannot be considered to be independent of one another, which is a fundamental assumption of the t procedures.

7.124. The mean is $\bar{x} \doteq 26.8437$ cm, $s \doteq 18.3311$ cm, and the margin of error is $t^* \cdot s/\sqrt{584}$:

	df	t^*	Interval
Table D	100	1.984	$26.8437 \pm 1.5050 = 25.3387$ to 28.3486 cm
Software	583	1.9640	$26.8437 \pm 1.4898 = 25.3538$ to 28.3335 cm

The confidence interval is much narrower with the whole data set, largely because the standard error is about one-fourth what it was with a sample of size 40. The distribution of the 584 measurements is right-skewed (although not as much as the smaller sample). If we can view these trees as an SRS of similar stands—a fairly questionable assumption—the t procedures should be fairly reliable because of the large sample size. See the solution to Exercise 7.126 for an examination of the distribution.

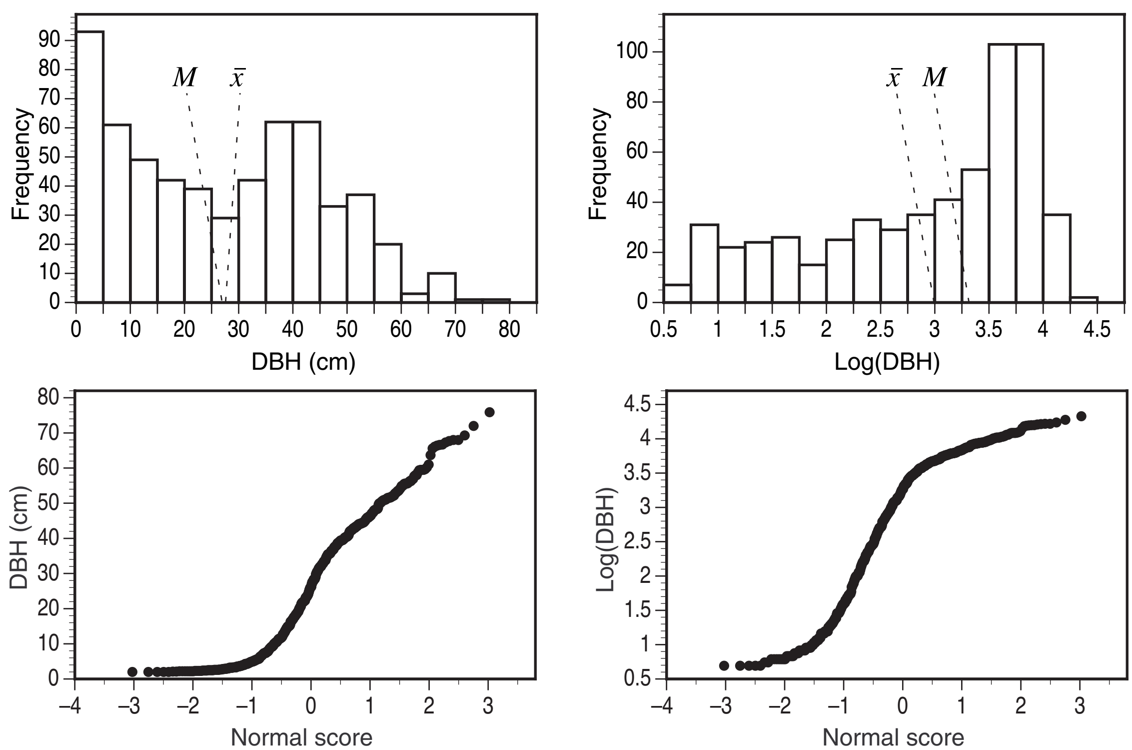
7.125. The tables on the following page contain summary statistics and 95% confidence intervals for the differences. For north/south differences, the test of H_0: $\mu_n = \mu_s$ gives $t = -7.15$ with df = 575.4 or 283; either way, $P < 0.0001$, so we reject H_0. For east/west differences, $t = -3.69$ with df = 472.7 or 230; either way, $P \doteq 0.0003$, so we reject H_0. The larger data set results in smaller standard errors (both are near 1.5, compared to about 4 in Exercises 7.81 and 7.82), meaning that t is larger and the margin of error is smaller.

	$\bar{x}$	s	n
North	21.7990	18.9230	300
South	32.1725	16.0763	284
East	24.5785	17.7315	353
West	30.3052	18.7264	231

	df	t^*	Confidence interval
N − S	575.4	1.9641	−13.2222 to −7.5248
	283	1.9684	−13.2285 to −7.5186
	100	1.984	−13.2511 to −7.4960
E − W	472.7	1.9650	−8.7764 to −2.6770
	230	1.9703	−8.7847 to −2.6687
	100	1.984	−8.8059 to −2.6475

7.126. The histograms and quantile plots are shown below, and the means and medians are given in the table on the right and are marked on the histograms. (The plots were created using natural logarithms; for common logs, the appearance would be roughly the same except for scale.) The transformed data does not look notably more Normal; it is left-skewed instead of right-skewed. The t procedures should be fairly dependable anyway because of the large sample size, but only if we can view the data as an SRS from some population.

	$\bar{x}$	M
Original	26.8437	26.15
Natural log	2.9138	3.2638
Common log	1.2654	1.4175



7.127. **(a)** This is a matched pairs design because at each of the 24 nests, the same mockingbird responded on each day. **(b)** The variance of the difference is approximately $s_1^2 + s_4^2 - 2\rho s_1 s_4 = 48.684$, so the standard deviation is 6.9774 m. **(c)** To test H_0: $\mu_1 = \mu_4$ versus H_a: $\mu_1 \neq \mu_4$, we have $t = \frac{15.1 - 6.1}{6.9774/\sqrt{24}} \doteq 6.319$ with df = 23, for which P is very small. **(d)** Assuming the correlation is the same ($\rho = 0.4$), the variance of the difference is approximately $s_1^2 + s_5^2 - 2\rho s_1 s_5 = 31.324$, so the standard deviation is 5.5968 m. To test H_0: $\mu_1 = \mu_5$ versus H_a: $\mu_1 \neq \mu_5$, we have $t = \frac{4.9 - 6.1}{5.5968/\sqrt{24}} \doteq -1.050$ with df = 23, for which $P \doteq 0.3045$. **(e)** The significant difference between day 1 and day 4 suggests that the mockingbirds altered their behavior when approached by the same person for four consecutive days; seemingly, the birds perceived an escalating threat. When approached by a new person on day 5, the response was not significantly different from day 1; this suggests that the birds saw the new person as less threatening than a return visit from the first person.

7.128. **(a)** In each case, we test H_0: $\mu_a = \mu_s$ versus H_a: $\mu_a \neq \mu_s$. The table below summarizes the values of $\text{SE}_D = \sqrt{s_a^2/n_a + s_s^2/n_s}$ and $t = \frac{\bar{x}_a - \bar{x}_s}{\text{SE}_D}$, and the two options for df and P. For the second option, P-values from Table D are obtained using df = 70. **(b)** We conclude that sedentary female high school students had significantly higher body fat, BMI, and calcium deficit than female athletes. The difference in milk consumption was not significant.

			Option 1		Option 2	
Group	SE_D	t	df	P	df	P
Body fat	1.0926	−6.315	140.1	$P < 0.0001$	79	$P < 0.001$
Body mass index	0.4109	−11.707	156.3	$P < 0.0001$	79	$P < 0.001$
Calcium deficit	71.2271	−3.979	143.7	$P \doteq 0.0001$	79	$P < 0.001$
Glasses of milk/day	0.2142	1.821	154.0	$P \doteq 0.0705$	79	$0.05 < P < 0.1$

7.129. The mean and standard deviation of the 25 numbers are $\bar{x} = 78.32\%$ and $s \doteq 33.3563\%$, so the standard error is $\text{SE}_{\bar{x}} \doteq 6.6713\%$. For df = 24, Table D gives $t^* = 2.064$, so the 95% confidence interval is $\bar{x} \pm 13.7695\% = 64.5505\%$ to 92.0895% (with software, $t^* = 2.0639$ and the interval is $\bar{x} \pm 13.7688\% = 64.5512\%$ to 92.0888%). This seems to support the retailer's claim: The original supplier's price was higher between 65% to 93% of the time.

7.130. **(a)** We are interested in weight change; the pairs are the "before" and "after" measurements. **(b)** The mean weight change was a loss. The exact amount lost is not specified, but it was large enough so that it would rarely happen by chance for an ineffective weight-loss program. **(c)** Comparing to a $t(40)$ distribution in Table D, we find $P < 0.0005$ for a one-sided alternative ($P < 0.0010$ for a two-sided alternative). Software reveals that it is even smaller than that: about 0.000013 (or 0.000026 for a two-sided alternative).

7.131. Back-to-back stemplots below. The distributions appear similar; the most striking difference is the relatively large number of boys with low GPAs. Testing the difference in GPAs (H_0: $\mu_b = \mu_g$; H_a: $\mu_b < \mu_g$), we obtain $\text{SE}_D \doteq 0.4582$ and $t = -0.91$, which is not significant, regardless of whether we use df $= 74.9$ ($P = 0.1839$) or 30 ($0.15 < P < 0.20$). The 95% confidence interval for the difference $\mu_b - \mu_g$ in GPAs is shown in the second table on the right.

		GPA		IQ	
	n	$\bar{x}$	s	$\bar{x}$	s
Boys	47	7.2816	2.3190	110.96	12.121
Girls	31	7.6966	1.7208	105.84	14.271

	df	t^*	Confidence interval
GPA	74.9	1.9922	−1.3277 to 0.4979
	30	2.042	−1.3505 to 0.5207
IQ	56.9	2.0025	−1.1167 to 11.3542
	30	2.042	−1.2397 to 11.4772

For the difference in IQs, we find $\text{SE}_D \doteq 3.1138$. With the same hypotheses as before, we find $t = 1.64$—fairly strong evidence, but not quite significant at the 5% level: $P = 0.0528$ (df $= 56.9$) or $0.05 < P < 0.10$ (df $= 30$). The 95% confidence interval for the difference $\mu_b - \mu_g$ in IQs is shown in the second table on the right.

GPA:

Girls		Boys
	0	5
	1	7
	2	4
4	3	689
7	4	068
952	5	0
4200	6	019
988855432	7	1124556666899
998731	8	001112238
95530	9	1113445567
17	10	57

IQ:

Girls		Boys
42	7	
	7	79
	8	
96	8	
31	9	03
86	9	77
433320	10	0234
875	10	556667779
44422211	11	00001123334
98	11	556899
0	12	03344
8	12	67788
20	13	
	13	6

7.132. The median self-concept score is 59.5. A back-to-back stemplot (below) suggests that high self-concept students have a higher mean GPA and IQ. Testing H_0: $\mu_{\text{low}} = \mu_{\text{high}}$ versus H_a: $\mu_{\text{low}} \neq \mu_{\text{high}}$ for GPAs leads to $\text{SE}_D \doteq 0.4167$ and $t = 4.92$, which is quite significant ($P < 0.0005$ regardless of df). The confidence interval for the difference $\mu_{\text{high}} - \mu_{\text{low}}$ in GPAs is shown in the second table on the right.

		GPA		IQ	
	n	$\bar{x}$	s	$\bar{x}$	s
High SC	39	8.4723	1.3576	114.23	10.381
Low SC	39	6.4208	2.2203	103.62	13.636

	df	t^*	Confidence interval
GPA	62.9	1.9984	1.2188 to 2.8843
	38	2.0244	1.2079 to 2.8951
	30	2.042	1.2006 to 2.9025
IQ	70.9	1.9940	5.1436 to 16.0871
	38	2.0244	5.0601 to 16.1707
	30	2.042	5.0118 to 16.2190

For the difference in IQs, we find $\text{SE}_D \doteq 2.7442$. With the same hypotheses as before, we find $t = 3.87$, which is quite significant ($P < 0.0006$ regardless of df). The confidence interval for the difference $\mu_{\text{high}} - \mu_{\text{low}}$ in IQs is shown in the second table on the right.

In summary, both differences are significant; with 95% confidence, high self-concept students have a mean GPA that is 1.2 to 2.9 points higher, and their mean IQ is 5 to 16 points higher.

GPA:

Low SC		High SC
5	0	
7	1	
4	2	
864	3	9
8760	4	
9520	5	
94000	6	12
986655421	7	1234556688899
887321100	8	112399
97	9	0111334455556
	10	1577

IQ:

Low SC		High SC
42	7	
97	7	
	8	
96	8	
310	9	3
776	9	8
44300	10	22333
9877766	10	55567
44432100	11	00111222334
99985	11	568
4	12	00334
7	12	67888
	13	02
	13	6

7.133. It is reasonable to have a prior belief that people who evacuated their pets would score higher, so we test H_0: $\mu_1 = \mu_2$ versus H_a: $\mu_1 > \mu_2$. We find $\text{SE}_D \doteq 0.4630$ and $t = 3.65$, which gives $P < 0.0005$ no matter how we choose degrees of freedom (115 or 237.0). As one might suspect, people who evacuated their pets have a higher mean score.

df	t^*	Confidence interval
237.0	1.9700	0.7779 to 2.6021
115	1.9808	0.7729 to 2.6071
100	1.984	0.7714 to 2.6086

One might also compute a 95% confidence interval for the difference; these are given in the table.

	n	Calories $\bar{x}$	Calories s	Alcohol $\bar{x}$	Alcohol s
Drivers	98	2821	435.58	0.24	0.5940
Conductors	83	2844	437.30	0.39	1.0021

df	t^*	Confidence interval
128.4	2.6146	−0.4776 to 0.1776
82	2.6371	−0.4804 to 0.1804
80	2.639	−0.4807 to 0.1807

7.134. (a) "se" is standard error (of the mean). To find s, multiply the standard error by $\sqrt{n}$. **(b)** No: We test H_0: $\mu_d = \mu_c$ versus H_a: $\mu_d < \mu_c$, for which $\text{SE}_D \doteq 65.1153$ and $t \doteq -0.3532$, so $P = 0.3623$ (df $= 173.9$) or 0.3625 (df $= 82$)—in either case, there is little evidence against H_0. **(c)** The evidence is not very significant: To test H_0: $\mu_d = \mu_c$ versus H_a: $\mu_d \neq \mu_c$, $\text{SE}_D \doteq 0.1253$, $t \doteq -1.1971$, for which $P = 0.2335$ (df $= 128.4$) or 0.2348 (df $= 82$). **(d)** The 95% confidence interval is $0.39 \pm t^*(0.11)$. With Table D, $t^* = 1.990$ (df $= 80$) and the interval is 0.1711 to 0.6089 g; with software, $t^* = 1.9893$ (df $= 82$) and the interval is 0.1712 to 0.6088 g. **(e)** The 99% confidence interval is $(0.24 - 0.39) \pm t^*\sqrt{0.06^2 + 0.11^2}$; see the table.

7.135. The similarity of the sample standard deviations suggests that the population standard deviations are likely to be similar. The pooled standard deviation is $s_p \doteq 436.368$ and $t \doteq -0.3533$, so $P = 0.3621$ (df $= 179$)—still not significant.

7.136. (a) The sample sizes (98 and 83) are quite large, so the t test should be reasonably safe (provided there are no extreme outliers). **(b)** Large samples do not make the F test more reliable when the underlying distributions are skewed, so it should not be used.

7.137. No: What we have is nothing like an SRS of the population of school corporations.

7.138. **(a)** Back-to-back stemplot below; summary statistics on the right. With a pooled standard deviation, $s_p \doteq 83.6388$, $t \doteq 4.00$ with df $= 222$, so $P < 0.0001$. Without pooling, $\text{SE}_D \doteq 11.6508$, $t \doteq 4.01$ with df $= 162.2$, and again $P < 0.0001$ (or, with df $= 78$, we conclude that $P < 0.0005$). The test for equality of standard deviations gives $F \doteq 1.03$ with df 144 and 78; the P-value is 0.9114, so the pooled procedure should be appropriate. In either case, we conclude that male mean SATM scores are higher than female mean SATM scores. Options for a 95% confidence interval for the male − female difference are given in the table below. **(b)** With a pooled standard deviation, $s_p \doteq 92.6348$, $t \doteq 0.94$ with df $= 222$, so $P = 0.3485$. Without pooling, $\text{SE}_D \doteq 12.7485$, $t \doteq 0.95$ with df $= 162.2$, so $P = 0.3410$ (or, with df $= 78$, $P = 0.3426$). The test for equality of standard deviations gives $F \doteq 1.11$ with df 144 and 78; the P-value is 0.6033, so the pooled procedure should be appropriate. In either case, we cannot see a difference between male and female mean SATV scores. Options for a 95% confidence interval for the male − female difference are given in the table below. **(c)** The results may generalize fairly well to students in different years, less well to students at other schools, and probably not very well to college students in general.

		SATM		SATV	
	n	$\bar{x}$	s	$\bar{x}$	s
Men	145	611.772	84.0206	508.841	94.3485
Women	79	565.025	82.9294	496.671	89.3849

df	t^*	Confidence interval for SATM $\mu_M - \mu_F$
162.3	1.9747	23.7402 to 69.7538
78	1.9908	23.5521 to 69.9419
70	1.994	23.5154 to 69.9786
222	1.9707	23.6978 to 69.7962
100	1.984	23.5423 to 69.9517

df	t^*	Confidence interval for SATV $\mu_M - \mu_F$
167.9	1.9742	−12.9981 to 37.3381
78	1.9908	−13.2104 to 37.5504
70	1.994	−13.2506 to 37.5906
222	1.9707	−13.3584 to 37.6984
100	1.984	−13.5306 to 37.8706

Men's SATM		Women's SATM
	3	0
	3	5
400	4	1334
99999888776	4	56777888999
444443333222222111000	5	0111123334
99999988888877766555555555	5	5555555666777777788899999
44444444443333333222222221110000000	6	00011222233334444
9999999877777766666655555	6	55555789
3222211100000	7	1124
77766655555	7	
0	8	

Men's SATV		Women's SATV
98	2	9
4322	3	33
9999988766	3	55669
44444444433211110000 0	4	0122223333444
999998888888888887777766666655 5	4	5666666777777888888899999
44443333322222111000000000000 0	5	01111122334
99888877777766666655 5	5	566777777889
433333321111000000	6	0000
9987775	6	668
420	7	00
6	7	5

	n	$\bar{x}$	s
Basal	22	41.0455	5.6356
DRTA	22	46.7273	7.3884
Strat	22	44.2727	5.7668

7.139. (a) We test H_0: $\mu_B = \mu_D$ versus H_a: $\mu_B < \mu_D$. Pooling might be appropriate for this problem, in which case $s_p \doteq 6.5707$. Whether or not we pool, $\text{SE}_D \doteq 1.9811$ and $t \doteq 2.87$ with df $= 42$ (pooled), 39.3, or 21, so $P = 0.0032$, or 0.0033, or 0.0046. We conclude that the mean score using DRTA is higher than the mean score with the Basal method. The difference in the average scores is 5.68; options for a 95% confidence interval for the difference $\mu_D - \mu_B$ are given in the table below. **(b)** We test H_0: $\mu_B = \mu_S$ versus H_a: $\mu_B < \mu_S$. If we pool, $s_p \doteq 5.7015$. Whether or not we pool, $\text{SE}_D \doteq 1.7191$ and $t \doteq 1.88$ with df $= 42$, 42.0, or 21, so $P = 0.0337$, or 0.0337, or 0.0372. We conclude that the mean score using Strat is higher than the Basal mean score. The difference in the average scores is 3.23; options for a 95% confidence interval for the difference $\mu_S - \mu_B$ are given in the table below.

df	t^*	Confidence interval for $\mu_D - \mu_B$
39.3	2.0223	1.6754 to 9.6882
21	2.0796	1.5618 to 9.8018
21	2.080	1.5610 to 9.8026
42	2.0181	1.6837 to 9.6799
40	2.021	1.6779 to 9.6857

df	t^*	Confidence interval for $\mu_S - \mu_B$
42.0	2.0181	−0.2420 to 6.6966
21	2.0796	−0.3477 to 6.8023
21	2.080	−0.3484 to 6.8030
42	2.0181	−0.2420 to 6.6965
40	2.021	−0.2470 to 6.7015

7.140. For testing $\mu_1 = \mu_2$ against a two-sided alternative, we would reject H_0 if $t = \frac{\bar{x}_1 - \bar{x}_2}{\text{SE}_D}$ is greater (in absolute value) than t^*, where $\text{SE}_D = 2.5\sqrt{2/n}$. (Rather than determining t^* for each considered sample size, we might use $t^* \doteq 2$.) We therefore want to choose n so that

$$P\left(\left|\frac{\bar{x}_1 - \bar{x}_2}{\text{SE}_D}\right| > t^*\right) = 1 - P\left(-t^* < \frac{\bar{x}_1 - \bar{x}_2}{\text{SE}_D} < t^*\right) = 0.90$$

when $\mu_1 - \mu_2 = 0.4$ inch. With $\delta = 0.4/\text{SE}_D \doteq 0.1131\sqrt{n}$, this means that

$$P\left(-t^* - \delta < Z < t^* - \delta\right) = 0.10$$

where Z has a $N(0, 1)$ distribution. By trial and error, we find that two samples of size 822 will do this. (This answer will vary slightly depending on what students do with t^* in the formula above.)

Note: *Determining the necessary sample size can be made a bit easier with some software. The output of the G•Power program below gives the total sample size as 1644 (i.e., two samples of size 822).*

G•Power output

```
A priori analysis for "t-Test (means)", two-tailed:
Alpha: 0.0500
Power (1-beta): 0.9000
Effect size "d": 0.1600
Total sample size: 1644
Actual power: 0.9001
Critical value: t(1642) = 1.9614
Delta: 3.2437
```

7.141. **(a)** The distributions can be compared using a back-to-back stemplot (shown), or two histograms, or side-by-side boxplots. Three-bedroom homes are right-skewed; four-bedroom homes are generally more expensive. The top two prices from the three-bedroom distribution qualify as outliers using the $1.5 \times IQR$ criterion. Boxplots are probably a poor choice for displaying the distributions because they leave out so much detail, but five-number summaries do illustrate that four-bedroom prices are higher at every level. Summary statistics (in units of \$1000) are given in the table below. **(b)** For testing H_0: $\mu_3 = \mu_4$ versus H_a: $\mu_3 \neq \mu_4$, we have $t \doteq -4.475$ with either df = 20.98 ($P \doteq 0.0002$) or df = 13 ($P < 0.001$). We reject H_0 and conclude that the mean prices are different (specifically, that 4BR houses are more expensive). **(c)** The one-sided alternative $\mu_3 < \mu_4$ could have been justified because it would be reasonable to expect that four-bedroom homes would be more expensive. **(d)** The 95% confidence interval for the difference $\mu_4 - \mu_3$ is about \$63,823 to \$174,642 (df = 20.97) or \$61,685 to \$176,779 (df = 13). **(e)** While the data were not gathered from an SRS, it seems that they should be a fair representation of three- and four-bedroom houses in West Lafayette. (Even so, the small sample sizes, together with the skewness and the outliers in the three-bedroom data, should make us cautious about the t procedures. Additionally, we might question independence in these data: When setting the asking price for a home, sellers are almost certainly influenced by the asking prices for similar homes on the market in the area.)

3BR		4BR
99987	0	
4432211100	1	4
8655	1	678
1	2	024
976	2	68
	3	223
	3	9
	4	2

	n	$\bar{x}$	s	Min	Q_1	M	Q_3	Max
3BR	23	147.561	61.741	79.5	100.0	129.9	164.9	295.0
4BR	14	266.793	87.275	149.9	189.0	259.9	320.0	429.9

Chapter 8 Solutions

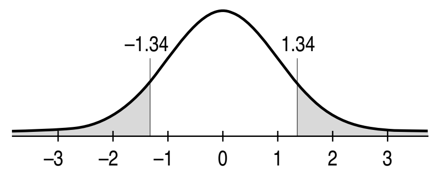
8.1. **(a)** $n = 760$ banks. **(b)** $X = 283$ banks expected to acquire another bank. **(c)** $\hat{p} = \frac{283}{760} \doteq 0.3724$.

8.2. **(a)** $n = 1063$ adults played video games. **(b)** $X = 223$ of those adults play daily or almost daily. **(c)** $\hat{p} = \frac{223}{1063} \doteq 0.2098$.

8.3. **(a)** With $\hat{p} \doteq 0.3724$, $\text{SE}_{\hat{p}} = \sqrt{\hat{p}(1-\hat{p})/760} \doteq 0.01754$. **(b)** The 95% confidence interval is $\hat{p} \pm 1.96\,\text{SE}_{\hat{p}} = 0.3724 \pm 0.0344$. **(c)** The interval is 33.8% to 40.7%.

8.4. **(a)** With $\hat{p} \doteq 0.2098$, $\text{SE}_{\hat{p}} = \sqrt{\hat{p}(1-\hat{p})/1063} \doteq 0.01249$. **(b)** The 95% confidence interval is $\hat{p} \pm 1.96\,\text{SE}_{\hat{p}} = 0.2098 \pm 0.0245$. **(c)** The interval is 18.5% to 23.4%.

8.5. For $z = 1.34$, the two-sided P-value is the area under a standard Normal curve above 1.34 and below -1.34.



8.6. The 95% confidence interval is 44% to 86%. (Student opinions of what qualifies as "appropriate" rounding might vary.) This is given directly in the Minitab output; the SAS output gives a confidence interval for the complementary proportion.

The confidence interval in consistent with the result of the significance test, but is more informative in that it gives a range of values for the true proportion.

8.7. The sample proportion is $\hat{p} = \frac{15}{20} = 0.75$. To test H_0: $p = 0.5$ versus H_a: $p \neq 0.5$, the appropriate standard error is $\sigma_{\hat{p}} = \sqrt{p_0(1-p_0)/20} \doteq 0.1118$, and the test statistic is $z = (\hat{p} - p_0)/\sigma_{\hat{p}} \doteq \frac{0.25}{0.1118} \doteq 2.24$. The two-sided P-value is 0.0250 (Table A) or 0.0253 (software), so this result is significant at the 5% level.

8.8. With $n = 40$ and $\hat{p} = 0.65$, the standard error for the significance test is $\sigma_{\hat{p}} = \sqrt{p_0(1-p_0)/40} \doteq 0.0791$, and the test statistic is $z = (\hat{p} - p_0)/\sigma_{\hat{p}} \doteq \frac{0.15}{0.0791} \doteq 1.90$. The two-sided P-value is 0.0574 (Table A) or 0.0578 (software)—not quite significant at the 5% level, but stronger evidence than the result with $n = 20$.

8.9. **(a)** To test H_0: $p = 0.5$ versus H_a: $p \neq 0.5$ with $\hat{p} = 0.35$, the test statistic is

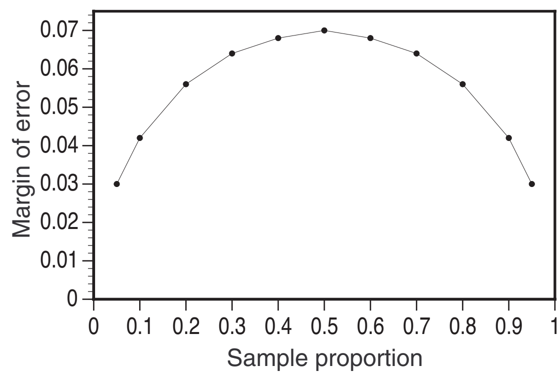
$$z = \frac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}} \doteq \frac{-0.15}{0.1118} \doteq -1.34$$

This is the opposite of the value of z given in Example 8.4, and the two-sided P-value is the same: 0.1802 (or 0.1797 with software). **(b)** The standard error for a confidence interval is $\text{SE}_{\hat{p}} = \sqrt{\hat{p}(1-\hat{p})/20} \doteq 0.1067$, so the 95% confidence interval is $0.35 \pm 0.2090 = 0.1410$ to 0.5590. This is the complement of the interval shown in the Minitab output in Figure 8.2.

8.10. We can achieve that margin of error with 90% confidence with a smaller sample. With $p^* = 0.5$ (as in Example 8.5), we compute $n = \left(\frac{1.645}{(2)(0.03)}\right)^2 \doteq 751.67$, so we need a sample of 752 students.

8.11. The plot is symmetric about 0.5, where it has its maximum. (The maximum margin of error always occurs at $\hat{p} = 0.5$, but the size of the maximum error depends on the sample size.)

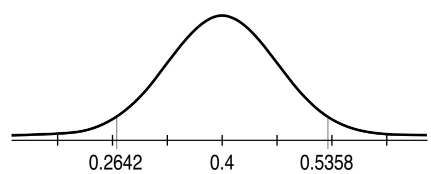
Note: *The first printing of the text asked students to plot sample proportion ($\hat{p}$) versus the margin of error (m), rather than m versus $\hat{p}$. Because $\hat{p}$ is the explanatory variable, the latter is more natural.*



8.12. **(a)** H_0 should refer to p (the population proportion), not $\hat{p}$ (the sample proportion). **(b)** Use Normal distributions (and a z test statistic) for significance tests involving proportions. **(c)** The margin of error equals z^* times standard error; for 95% confidence, we would have $z^* = 1.96$.

8.13. **(a)** Margin of error only accounts for random sampling error. **(b)** P-values measure the strength of the evidence against H_0, not the probability of it being true. **(c)** The confidence level cannot exceed 100%. (In practical terms, the confidence level must be *less than* 100%.)

8.14. **(a)** The mean is $\mu = p = 0.4$ and the standard deviation is $\sigma = \sqrt{p(1-p)/n} = \sqrt{0.0048} \doteq 0.06928$. **(b)** Normal curve on the right. The tick marks are σ (about 0.07) units apart. **(c)** p^* should be either $1.96\sigma \doteq 0.1358$ or $2\sigma \doteq 0.1386$, so the points marked on the curve should be either 0.2642 and 0.5358 or 0.2614 and 0.5386.



8.15. The sample proportion is $\hat{p} = \frac{3274}{5000} \doteq 0.6548$, the standard error is $\mathrm{SE}_{\hat{p}} \doteq 0.00672$, and the 95% confidence interval is $0.6548 \pm 0.0132 = 0.6416$ to 0.6680.

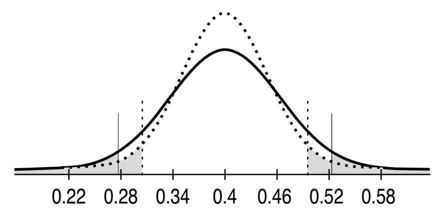
8.16. **(a)** With $\hat{p} = 0.16$, we have $\mathrm{SE}_{\hat{p}} \doteq 0.00518$, so the 95% confidence interval is $\hat{p} \pm 1.96\,\mathrm{SE}_{\hat{p}} = 0.16 \pm 0.01016 \doteq 0.1498$ to 0.1702. **(b)** With 95% confidence, the percent of this group who would prefer a health care professional as a marriage partner is about $16 \pm 1\%$, or 15% and 17%.

8.17. **(a)** $\text{SE}_{\hat{p}}$ depends on n, which is some number less than 7061. Without that number, we do not know the margin of error $z^*\text{SE}_{\hat{p}}$. **(b)** The number who expect to begin playing an instrument is 67% of half of 7061, or $(0.67)(0.5)(7061) \doteq 2365$ players. **(c)** Taking $n = (0.5)(7061) = 3530.5$, the 99% confidence interval is $\hat{p} \pm 2.576\,\text{SE}_{\hat{p}} = 0.67 \pm 0.02038 \doteq$ 0.6496 to 0.6904. **(d)** We do not know the sampling methods used, which might make these methods unreliable.

Note: *Even though n must be an integer in reality, it is not necessary to round n in part (c); the confidence interval formula works fine when n is not a whole number.*

8.18. **(a)** With $n = (0.25)(7061) = 1765.25$, the 99% confidence interval is $\hat{p} \pm 2.576\,\text{SE}_{\hat{p}} = 0.67 \pm 0.02883 \doteq 0.6412$ to 0.6988. **(b)** With $n = (0.75)(7061) = 5295.75$, the 99% confidence interval is $0.67 \pm 0.01664 \doteq 0.6534$ to 0.6866. **(c)** The margin of error depends on the (unknown) sample size, and varies quite a bit among the three scenarios.

8.19. If $\hat{p}$ has (approximately) a $N(p_0, \sigma)$ distribution under H_0, we reject H_0 (at the $\alpha = 0.05$ level) if $\hat{p}$ falls outside the range $p_0 \pm 1.96\sigma$. **(a)** If H_0 is true and $n = 60$, then $\sigma = \sqrt{(0.4)(0.6)/60} \doteq 0.06325$, so we reject H_0 when $\hat{p}$ is outside the range 0.2760 to 0.5240. Because $\hat{p} = \frac{x}{60}$, this corresponds to: Reject H_0 if $\hat{p} \le 0.2667$ ($x \le 15$) or $\hat{p} \ge 0.5333$ ($x \ge 32$). **(b)** If H_0 is true and $n = 100$, then $\sigma = \sqrt{(0.4)(0.6)/100} \doteq 0.04899$, so we reject H_0 when $\hat{p}$ is outside the range 0.3040 to 0.4960. This corresponds to: Reject H_0 if $\hat{p} \le 0.30$ ($x \le 30$) or $\hat{p} \ge 0.50$ ($x \ge 50$). **(c)** Shown on the right is one possible sketch, with the two Normal curves drawn on the same scale; the dashed curve and rejection cutoffs is for $n = 100$. (Most students will likely not realize that when σ is smaller, the curve must be taller to compensate for the decreased width.) With a larger sample size, smaller values of $|\hat{p} - 0.4|$ lead to the rejection of H_0.



8.20. **(a)** For 99% confidence, the margin of error is $(2.576)(0.00122) \doteq 0.00314$. **(b)** All of these facts suggest possible sources of error; for example, students at two-year colleges are not represented, nor are students at institutions that do not wish to pay the fee. (Even though the fee is scaled for institution size, larger institutions can more easily absorb it.) None of these potential errors are covered by the margin of error found in part (a), which only accounts for random sampling error.

8.21. **(a)** About $(0.42)(159{,}949) \doteq 67{,}179$ students plan to study abroad. **(b)** $\text{SE}_{\hat{p}} \doteq 0.00123$, the margin of error is $2.576\,\text{SE}_{\hat{p}} \doteq 0.00318$, and the 99% confidence interval is 0.4168 to 0.4232.

8.22. With $\hat{p} = \frac{1087}{1430} \doteq 0.7601$, we have $\text{SE}_{\hat{p}} \doteq 0.0113$, and the 95% confidence interval is $\hat{p} \pm 1.96\,\text{SE}_{\hat{p}} = 0.7601 \pm 0.0221 \doteq 0.7380$ to 0.7823.

8.23. With $\hat{p} = 0.43$, we have $\text{SE}_{\hat{p}} \doteq 0.0131$, and the 95% confidence interval is $\hat{p} \pm 1.96\,\text{SE}_{\hat{p}} = 0.43 \pm 0.0257 \doteq 0.4043$ to 0.4557.

8.24. **(a)** A 99% confidence interval would be wider: We need a larger margin of error (by a factor of 2.576/1.96) in order to be more confident that we have included p. The 99% confidence interval is 0.3963 to 0.4637. **(b)** A 90% confidence interval would be narrower (by a factor of 1.645/1.96). The 90% confidence interval is 0.4085 to 0.4515.

8.25. **(a)** $\mathrm{SE}_{\hat{p}} = \sqrt{(0.87)(0.13)/430,000} \doteq 0.0005129$. For 99% confidence, the margin of error is $2.576\,\mathrm{SE}_{\hat{p}} \doteq 0.001321$. **(b)** One source of error is indicated by the wide variation in response rates: We cannot assume that the statements of respondents represent the opinions of nonrespondents. The effect of the participation fee is harder to predict, but one possible impact is on the types of institutions that participate in the survey: Even though the fee is scaled for institution size, larger institutions can more easily absorb it. These other sources of error are much more significant than sampling error, which is the only error accounted for in the margin of error from part (a).

8.26. **(a)** The standard error is $\mathrm{SE}_{\hat{p}} = \sqrt{(0.69)(0.31)/1048} \doteq 0.01429$, so the margin of error for 95% confidence is $1.96\,\mathrm{SE}_{\hat{p}} \doteq 0.02800$ and the interval is 0.6620 to 0.7180. **(b)** To test H_0: $p = 0.79$ versus H_a: $p < 0.79$, the standard error is $\sigma_{\hat{p}} = \sqrt{(0.79)(0.21)/1048} \doteq 0.01258$ and the test statistic is $z = \frac{0.69-0.79}{0.01258} \doteq -7.95$. This is very strong evidence against H_0 ($P < 0.00005$).

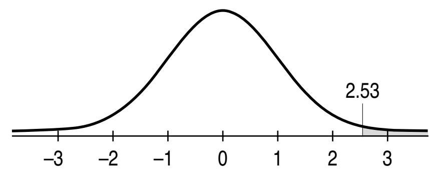
8.27. **(a)** The standard error is $\mathrm{SE}_{\hat{p}} = \sqrt{(0.38)(0.62)/1048} \doteq 0.01499$, so the margin of error for 95% confidence is $1.96\,\mathrm{SE}_{\hat{p}} \doteq 0.02939$ and the interval is 0.3506 to 0.4094. **(b)** Yes; some respondents might not admit to such behavior. The true frequency of such actions might be higher than this survey suggests.

8.28. **(a)** $\hat{p} = \frac{9054}{24{,}142} \doteq 0.3750$. **(b)** The standard error is $\mathrm{SE}_{\hat{p}} = \sqrt{\hat{p}(1-\hat{p})/24{,}142} \doteq 0.003116$, so the margin of error for 95% confidence is $1.96\,\mathrm{SE}_{\hat{p}} \doteq 0.00611$ and the interval is 0.3689 to 0.3811. **(c)** The nonresponse rate was $\frac{37{,}328-24{,}142}{37{,}328} \doteq 0.3532$—about 35%. We have no way of knowing if cheating is more or less prevalent among nonrespondents; this weakens the conclusions we can draw from this survey.

8.29. **(a)** $\hat{p} = \frac{390}{1191} \doteq 0.3275$. The standard error is $\mathrm{SE}_{\hat{p}} = \sqrt{\hat{p}(1-\hat{p})/1191} \doteq 0.01360$, so the margin of error for 95% confidence is $1.96\,\mathrm{SE}_{\hat{p}} \doteq 0.02665$ and the interval is 0.3008 to 0.3541. **(b)** Speakers and listeners probably perceive sermon length differently (just as, say, students and lecturers have different perceptions of the length of a class period).

8.30. A 90% confidence interval would be narrower: The margin of error will be smaller (by a factor of 1.645/2.576) if we are willing to be less confident that we have included p. The 90% confidence interval is 0.6570 to 0.6830—narrower than the 99% confidence interval (0.6496 to 0.6904) from Exercise 8.17.

8.31. Recall the rule of thumb from Chapter 5: Use the Normal approximation if $np \geq 10$ and $n(1-p) \geq 10$. We use p_0 (the value specified in H_0) to make our decision.
(a) No: $np_0 = 6$. **(b)** Yes: $np_0 = 18$ and $n(1-p_0) = 12$. **(c)** Yes: $np_0 = n(1-p_0) = 50$. **(d)** No: $np_0 = 2$.



8.32. (a) Because we have defined p as the proportion who prefer fresh-brewed coffee, we should compute $\hat{p} = \frac{28}{40} = 0.7$. To test H_0: $p = 0.5$ versus H_a: $p > 0.5$, the standard error is $\sigma_{\hat{p}} = \sqrt{(0.5)(0.5)/40} \doteq 0.07906$, and the test statistic is $z = \frac{0.7-0.5}{0.07906} \doteq 2.53$. The P-value is 0.0057. **(b)** Curve on the right. **(c)** The result is significant at the 5% level, so we reject H_0 and conclude that a majority of people prefer fresh-brewed coffee.

8.33. With $\hat{p} = 0.69$, $\text{SE}_{\hat{p}} \doteq 0.02830$ and the 95% confidence interval is 0.6345 to 0.7455.

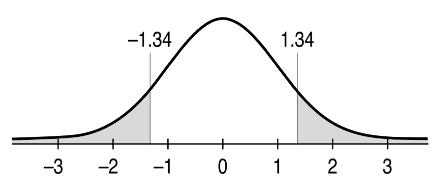
8.34. With $\hat{p} = 0.583$, $\text{SE}_{\hat{p}} \doteq 0.03023$ and the 95% confidence interval is 0.5237 to 0.6423.

8.35. We estimate $\hat{p} = \frac{594}{2533} \doteq 0.2345$, $\text{SE}_{\hat{p}} \doteq 0.00842$, and the 95% confidence interval is 0.2180 to 0.2510.

8.36. (a) We estimate $\hat{p} = \frac{1434}{2533} \doteq 0.5661$, $\text{SE}_{\hat{p}} \doteq 0.00985$, and the 95% confidence interval is 0.5468 to 0.5854. **(b)** Pride or embarrassment might lead respondents to claim that their income was above \$25,000 even if it were not. Consequently, it would not be surprising if the true proportion p were lower than the estimate $\hat{p}$. (There may also be some who would understate their income, out of humility or mistrust of the interviewer. While this would seem to have less of an impact, it makes it difficult to anticipate the overall effect of untruthful responses.) **(c)** Respondents would have little reason to lie about pet ownership; the few that might lie about it would have little impact on our conclusions. The number of untruthful responses about income is likely to be much larger and have a greater impact.

8.37. We estimate $\hat{p} = \frac{110}{125} = 0.88$, $\text{SE}_{\hat{p}} \doteq 0.02907$, and the 95% confidence interval is 0.8230 to 0.9370.

8.38. (a) $\hat{p} = \frac{542}{1711} \doteq 0.3168$; about 31.7% of bicyclists aged 15 or older killed between 1987 and 1991 had alcohol in their systems at the time of the accident. **(b)** $\text{SE}_{\hat{p}} = \sqrt{\hat{p}(1-\hat{p})/1711} \doteq 0.01125$; the 99% confidence interval is $\hat{p} \pm 2.576\,\text{SE}_{\hat{p}} = 0.2878$ to 0.3457. **(c)** No: We do not know, for example, what percent of cyclists who were *not* involved in fatal accidents had alcohol in their systems. **(d)** $\hat{p} = \frac{386}{1711} \doteq 0.2256$, $\text{SE}_{\hat{p}} \doteq 0.01010$, and the 99% confidence interval is 0.1996 to 0.2516.



8.39. (a) For testing H_0: $p = 0.5$ versus H_a: $p \neq 0.5$, we have $\hat{p} = \frac{5067}{10000} = 0.5067$ and $\sigma_{\hat{p}} = \sqrt{(0.5)(0.5)/10000} = 0.005$, so $z = \frac{0.0067}{0.005} = 1.34$, for which $P = 0.1802$. This is not significant at $\alpha = 0.05$ (or even $\alpha = 0.10$). **(b)** $\text{SE}_{\hat{p}} = \sqrt{\hat{p}(1-\hat{p})/10000} \doteq 0.005$, so the 95% confidence interval is $0.5067 \pm (1.96)(0.005)$, or 0.4969 to 0.5165.

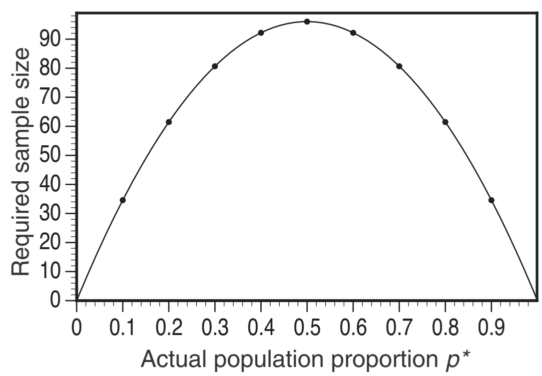
8.40. With no prior knowledge of the value of p (the proportion of "Yes" responses), take $p^* = 0.5$: $n = \left(\frac{1.96}{2(0.15)}\right)^2 \doteq 42.7$—use $n = 43$.

8.41. As a quick estimate, we can observe that to cut the margin of error in half, we must quadruple the sample size, from 43 to 172. Using the sample-size formula, we find $n = \left(\frac{1.96}{2(0.075)}\right)^2 \doteq 170.7$—use $n = 171$. (The difference in the two answers is due to rounding.)

8.42. Using $p^* = 0.25$ (based on previous surveys), we compute $n = \left(\frac{1.96}{0.1}\right)^2 (0.25)(0.75) \doteq 72.03$, so we need a sample of 73 students.

8.43. The required sample sizes are found by computing $\left(\frac{1.96}{0.1}\right)^2 p^*(1-p^*) = 384.16p^*(1-p^*)$: To be sure that we meet our target margin of error, we should take the largest sample indicated: $n = 97$ or larger.

p^*	n	Rounded up
0.1	34.57	35
0.2	61.47	62
0.3	80.67	81
0.4	92.20	93
0.5	96.04	97
0.6	92.20	93
0.7	80.67	81
0.8	61.47	62
0.9	34.57	35



8.44. $n = \left(\frac{1.96}{0.02}\right)^2 (0.15)(0.85) = 1224.51$—use $n = 1225$.

8.45. With $p_1 = 0.4$, $n_1 = 25$, $p_2 = 0.5$, and $n_2 = 30$, the mean and standard deviation of the sampling distribution of $D = \hat{p}_1 - \hat{p}_2$ are

$$\mu_D = p_1 - p_2 = -0.1 \text{ and } \sigma_D = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} \doteq 0.1339$$

8.46. **(a)** With $p_1 = 0.4$, $n_1 = 100$, $p_2 = 0.5$, and $n_2 = 120$, the mean and standard deviation of the sampling distribution of $D = \hat{p}_1 - \hat{p}_2$ are

$$\mu_D = p_1 - p_2 = -0.1 \text{ and } \sigma_D = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} \doteq 0.0670$$

(b) μ_D is unchanged, with σ_D is halved.

8.47. **(a)** The means are $\mu_{\hat{p}_1} = p_1$ and $\mu_{\hat{p}_2} = p_2$. The standard deviations are

$$\sigma_{\hat{p}_1} = \sqrt{\frac{p_1(1-p_1)}{n_1}} \text{ and } \sigma_{\hat{p}_2} = \sqrt{\frac{p_2(1-p_2)}{n_2}}$$

(b) $\mu_D = \mu_{\hat{p}_1} - \mu_{\hat{p}_2} = p_1 - p_2$. **(c)** $\sigma_D^2 = \sigma_{\hat{p}_1}^2 + \sigma_{\hat{p}_2}^2 = \dfrac{p_1(1-p_1)}{n_1} + \dfrac{p_2(1-p_2)}{n_2}$.

8.48. With $\hat{p}_w = \frac{44}{100} = 0.44$ and $\hat{p}_m = \frac{79}{140} \doteq 0.5643$, we estimate the difference to be $\hat{p}_m - \hat{p}_w \doteq 0.1243$. The standard error of the difference is

$$\text{SE}_D = \sqrt{\frac{\hat{p}_w(1-\hat{p}_w)}{100} + \frac{\hat{p}_m(1-\hat{p}_m)}{140}} \doteq 0.06496$$

so the 95% confidence interval for $p_m - p_w$ is $0.1243 \pm (1.96)(0.06496) = -0.0030$ to 0.2516.

Note: *We followed the text's practice of subtracting the smaller proportion from the larger one, as described at the top of page 494.*

8.49. Let us call the proportions favoring Commercial B q_w and q_m. Our estimates of these proportions are the complements of those found in Exercise 8.48; for example, $\hat{q}_w = \frac{56}{100} = 0.56 = 1 - \hat{p}_w$. Consequently, the standard error of the difference $\hat{q}_w - \hat{q}_m$ is the same as that for $\hat{p}_m - \hat{p}_w$: $\text{SE}_D = \sqrt{\frac{\hat{q}_w(1-\hat{q}_w)}{100} + \frac{\hat{q}_m(1-\hat{q}_m)}{140}} \doteq 0.06496$. The margin of error is therefore also the same, and the 95% confidence interval for $q_w - q_m$ is $(\hat{q}_w - \hat{q}_m) \pm (1.96)(0.06496) = -0.0030$ to 0.2516.

Note: *As in the previous exercise, we followed the text's practice of subtracting the smaller proportion from the larger one.*

8.50. The pooled estimate of the proportion is $\hat{p} = \frac{44+79}{100+140} = 0.5125$. For testing H_0: $p_m = p_w$ versus H_a: $p_m \neq p_w$, we have $\text{SE}_{D_p} = \sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{100} + \frac{1}{140}\right)} \doteq 0.06544$ and the test statistic is $z = (\hat{p}_m - \hat{p}_w)/\text{SE}_{D_p} \doteq 1.90$, for which the two-sided P-value is 0.0576. This is not quite enough evidence to reject H_0 at the 5% level.

8.51. Because the sample proportions would tend to support the alternative hypothesis ($p_m > p_w$), the P-value is half as large ($P = 0.0288$), which would be enough to reject H_0 at the 5% level.

8.52. **(a)** The filled-in table is on the right. **(b)** The estimated difference is $\hat{p}_2 - \hat{p}_1 \doteq 0.1198$. **(c)** Large-sample methods should be appropriate, because we have large, independent samples from two populations. **(d)** With $\text{SE}_D \doteq 0.01105$, the 95% confidence interval is $0.1198 \pm 0.02167 \doteq 0.0981$ to 0.1415. **(e)** The estimated difference is about 12.0%, and the interval is about 9.8% to 14.1%. **(f)** It is hard to imagine why the months for each survey would affect the interpretation. (Of course, just because we cannot guess what the impact would be does not mean there is no impact.)

Population	Population proportion	Sample size	Count of successes	Sample proportion
1	p_1	2822	198	0.0702
2	p_2	1553	295	0.1900

8.53. For H_0: $p_1 = p_2$, the pooled estimate of the proportion is $\hat{p} = \frac{198+295}{2822+1553} \doteq 0.1127$. The standard error is $\text{SE}_{D_p} \doteq 0.00999$, and the test statistic is $z = \frac{0.1198}{0.00999} \doteq 11.99$. The alternative hypothesis was not specified in this exercise; for either $p_1 \neq p_2$ or $p_1 < p_2$, the P-value associated with $z = 11.99$ would be tiny. (For the alternative $p_1 > p_2$, P would be nearly 1, and we would not reject H_0; however, it is hard to imagine why we would suspect that podcast use had decreased from 2006 to 2008.)

8.54. **(a)** No; this is a ratio of *proportions,* not *people*. In addition, these are sample proportions, so they are only estimates of the population proportions. If the size of the population (Internet users) remained roughly constant from 2006 to 2008, we can say that about *about* 2.7 times as many people are downloading podcasts. **(b)** We are quite confident that the 2008 proportion exceeds the 2006 proportion by at least 0.098, so making the (extremely reasonable) assumption that the number of Internet users did not decrease, we are nearly certain that there are more people downloading podcasts.

8.55. **(a)** The filled-in table is on the right. The values of X_1 and X_2 are estimated as (0.54)(1063) and (0.89)(1064). **(b)** The estimated difference is $\hat{p}_2 - \hat{p}_1 \doteq 0.35$.

Population	Population proportion	Sample size	Count of successes	Sample proportion
1	p_1	1063	574	0.54
2	p_2	1064	947	0.89

(c) Large-sample methods should be appropriate, because we have large, independent samples from two populations. **(d)** With $\text{SE}_D \doteq 0.01805$, the 95% confidence interval is $0.35 \pm 0.03537 \doteq 0.3146$ to 0.3854. **(e)** The estimated difference is about 35%, and the interval is about 31.5% to 38.5%. **(f)** A possible concern is that adults were surveyed before Christmas, while teens were surveyed before and after Christmas. It might be that some of those teens may have received game consoles as gifts, but eventually grew tired of them.

8.56. The pooled estimate of the proportion is $\hat{p} = \frac{(0.54)(1063)+(0.89)(1064)}{1063+1064} \doteq 0.7151$. For testing H_0: $p_1 = p_1$ versus H_a: $p_1 \neq p_2$, we have $\text{SE}_{D_p} = \sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{1063} + \frac{1}{1064}\right)} \doteq 0.01957$ and the test statistic is $z = (\hat{p}_2 - \hat{p}_1)/\text{SE}_{D_p} \doteq 17.88$. The P-value is essentially 0; we have no doubt that the two proportions are different (specifically, that the teen proportion is higher).

8.57. **(a)** The filled-in table is on the right. The values of X_1 and X_2 are estimated as (0.73)(1063) and (0.76)(1064). **(b)** The estimated difference is $\hat{p}_2 - \hat{p}_1 \doteq 0.03$.

Population	Population proportion	Sample size	Count of successes	Sample proportion
1	p_1	1063	776	0.73
2	p_2	1064	809	0.76

(c) Large-sample methods should be appropriate, because we have large, independent samples from two populations. **(d)** With $\text{SE}_D \doteq 0.01889$, the 95% confidence interval is $0.03 \pm 0.03702 \doteq -0.0070$ to 0.0670. **(e)** The estimated difference is about 3%, and the interval is about −0.7% to 6.7%. **(f)** As in the solution to Exercise 8.55, a possible concern is that adults were surveyed before Christmas.

8.58. The pooled estimate of the proportion is $\hat{p} = \frac{(0.73)(1063)+(0.76)(1064)}{1063+1064} \doteq 0.7450$. For testing H_0: $p_1 = p_1$ versus H_a: $p_1 \neq p_2$, we have $\text{SE}_{D_p} = \sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{1063} + \frac{1}{1064}\right)} \doteq 0.01890$ and the test statistic is $z = (\hat{p}_2 - \hat{p}_1)/\text{SE}_{D_p} \doteq 1.59$. The P-value is 0.1118 (Table A) or 0.1125 (software); either way, there is not enough evidence to conclude that the proportions are different.

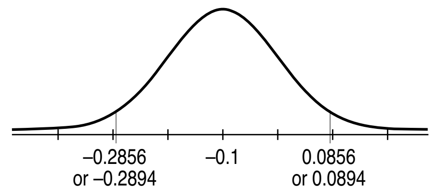
8.59. No; this procedure requires independent samples from different populations. We have one sample (of teens).

8.60. **(a)** The mean is

$$\mu_D = p_1 - p_2 = 0.4 - 0.5 = -0.1$$

and the standard deviation is

$$\sigma_D = \sqrt{\frac{p_1(1-p_1)}{50} + \frac{p_2(1-p_2)}{60}} \doteq 0.09469$$



(b) Normal curve on the right. **(c)** d should be either $1.96\sigma_D \doteq 0.1856$ or $2\sigma_D \doteq 0.1894$, so the points marked on the curve should be either -0.2856 and 0.0856 or -0.2894 and 0.0894.

Note: *Because this problem told us which population was "first" and which was "second," we did not follow the suggestion in the text to arrange them so that the population 1 had the larger proportion. Where necessary, we have done so in the other exercises.*

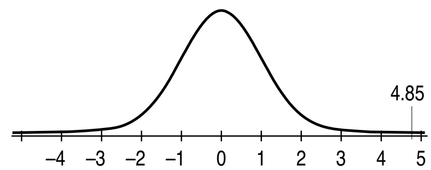
8.61. **(a)** H_0 should refer to p_1 and p_2 (population proportions) rather than $\hat{p}_1$ and $\hat{p}_2$ (sample proportions). **(b)** Knowing $\hat{p}_1 = \hat{p}_2$ does not tell us that the success counts are equal ($X_1 = X_2$) *unless* the sample sizes are equal ($n_1 = n_2$). **(c)** Confidence intervals only account for random sampling error.

8.62. **(a)** The mean of $D = \hat{p}_1 - \hat{p}_2$ is $\mu_D = p_1 - p_2 = 0.4 - 0.5 = -0.1$ (as before). The standard deviation is

$$\sigma_D = \sqrt{\frac{p_1(1-p_1)}{50} + \frac{p_2(1-p_2)}{1000}} \doteq 0.07106$$

(b) The mean of $\hat{p}_1 - 0.5$ is also -0.1; the standard deviation is the same as that of $\hat{p}_1$: $\sqrt{\frac{p_1(1-p_1)}{50}} \doteq 0.06928$. **(c)** The standard deviation of $\hat{p}_2$ is only 0.01581, so it will typically (95% of the time) differ from its mean (0.5) by no more than about 0.032. **(d)** If one sample is very large, that estimated proportion will be more accurate, and most of the variation in the difference will come from the variation in the other proportion.

8.63. Pet owners had the lower proportion of women, so we call them "population 2": $\hat{p}_2 = \frac{285}{595} \doteq 0.4790$. For non-pet owners, $\hat{p}_1 = \frac{1024}{1939} \doteq 0.5281$. $\text{SE}_D \doteq 0.02341$, so the 95% confidence interval is 0.0032 to 0.0950.



8.64. **(a)** Arranging the proportion so that population 1 has the larger proportion, we have $\hat{p}_1 = \frac{35}{165} \doteq 0.2121$ and $\hat{p}_2 = \frac{17}{283} \doteq 0.0601$. **(b)** $\hat{p}_1 - \hat{p}_2 \doteq 0.1521$ and the standard error (for constructing a confidence interval) is $\text{SE}_D \doteq 0.03482$. **(c)** The hypotheses are H_0: $p_1 = p_2$ versus H_a: $p_1 > p_2$. The alternative reflects the reasonable expectation that reducing pollution might decrease wheezing. **(d)** The pooled estimate of the proportion is $\hat{p} = \frac{17+35}{283+165} \doteq 0.1161$ and $\text{SE}_{D_p} \doteq 0.03137$, so $z = (\hat{p}_1 - \hat{p}_2)/\text{SE}_{D_p} \doteq 4.85$. The P-value is very small ($P < 0.0001$). **(e)** The 95% confidence interval, using the standard error from part (b), has margin of error $1.96\,\text{SE}_D \doteq 0.06824$: 0.0838 to 0.2203. The percent reporting improvement was between 8% and 22% higher for bypass residents. **(f)** There may be geographic factors (e.g., weather) or cultural factors (e.g., diet) that limit how much we can generalize the conclusions.

8.65. With equal sample sizes, the pooled estimate of the proportion is $\hat{p} = 0.255$, the average of $\hat{p}_1 = 0.29$ and $\hat{p}_2 = 0.22$. This can also be computed by taking $X_1 = (0.29)(1421) = 412.09$ and $X_2 = (0.22)(1421) = 312.62$, so $\hat{p} = (X_1 + X_2)/(1421 + 1421)$. The standard error for a significance test is $\text{SE}_{D_p} \doteq 0.01635$, and the test statistic is $z \doteq 4.28$ ($P < 0.0001$); we conclude that the proportions are different. The standard error for a confidence interval is $\text{SE}_D \doteq 0.01630$, and the 95% confidence interval is 0.0381 to 0.1019. The interval gives us an idea of how large the difference is: Music downloads dropped 4% to 10%.

8.66. The table below shows the results from the previous exercise, and those with different sample sizes. For part (*iii*), two answers are given, corresponding to the two ways one could interpret which is the "first sample size."

	n_1	n_2	$\hat{p}$	SE_{D_p}	z	SE_D	Confidence interval
8.65	1421	1421	0.255	0.01635	4.28	0.01630	0.0381 to 0.1019
(*i*)	1000	1000	0.255	0.01949	3.59	0.01943	0.0319 to 0.1081
(*ii*)	1600	1600	0.255	0.01541	4.54	0.01536	0.0399 to 0.1001
(*iii*)	1000	1600	0.2469	0.01738	4.03	0.01770	0.0353 to 0.1047
	1600	1000	0.2631	0.01775	3.94	0.01733	0.0360 to 0.1040

As one would expect, we see in (*i*) and (*ii*) that smaller samples result in smaller z (weaker evidence) and wider intervals, while larger samples have the reverse effect. The results of (*iii*) show that the effect of varying unequal sample sizes is more complicated.

8.67. **(a)** We find $\hat{p}_1 = \frac{73}{91} \doteq 0.8022$ and $\hat{p}_2 = \frac{75}{109} \doteq 0.6881$. For a confidence interval, $\text{SE}_D \doteq 0.06093$, so the 95% confidence interval for $p_1 - p_2$ is $(0.8022 - 0.6881) \pm (1.96)(0.06093) = -0.0053$ to 0.2335. **(b)** The question posed was, "Do high-tech companies tend to offer stock options more often than other companies?" Therefore, we test H_0: $p_1 = p_2$ versus H_a: $p_1 > p_2$. With $\hat{p}_1 \doteq 0.8022$, $\hat{p}_2 \doteq 0.6881$, and $\hat{p} = \frac{73+75}{91+109} = 0.74$, we find $\text{SE}_{D_p} \doteq 0.06229$, so $z = (\hat{p}_1 - \hat{p}_2)/\text{SE}_{D_p} \doteq 1.83$. This gives $P = 0.0336$. **(c)** We have fairly strong evidence that high-tech companies are more likely to offer stock options. However, the confidence interval tells us that the difference in proportions could be very small, or as large as 23%.

8.68. With $\hat{p}_{2002} \doteq 0.4780$ and $\hat{p}_{2004} \doteq 0.3750$, the standard error for a confidence interval is $SE_D \doteq 0.00550$. The 90% confidence interval for the difference $p_{2002} - p_{2004}$ is $(0.4780 - 0.3750) \pm 1.645 SE_D = 0.0939$ to 0.1120.

8.69. **(a)** $\hat{p}_f = \frac{48}{60} = 0.8$, so $SE_{\hat{p}} \doteq 0.05164$ for females. $\hat{p}_m = \frac{52}{132} = 0.\overline{39}$, so $SE_{\hat{p}} \doteq 0.04253$ for males. **(b)** $SE_D = \sqrt{0.05164^2 + 0.04253^2} \doteq 0.06690$, so the interval is $(\hat{p}_f - \hat{p}_m) \pm 1.645\, SE_D$, or 0.2960 to 0.5161. There is (with high confidence) a considerably higher percent of juvenile references to females than to males.

8.70. **(a)** $\hat{p}_1 = \frac{515}{1520} \doteq 0.3388$ for men, and $\hat{p}_2 = \frac{27}{191} \doteq 0.1414$ for women. $SE_D \doteq 0.02798$, so the 95% confidence interval for $p_1 - p_2$ is 0.1426 to 0.2523. **(b)** The female contribution is larger because the sample size for women is much smaller. (Specifically, $\hat{p}_1(1 - \hat{p}_1)/n_1 \doteq 0.0001474$, while $\hat{p}_2(1 - \hat{p}_2)/n_2 \doteq 0.0006355$.) Note that if the sample sizes had been similar, the male contribution would have been larger (assuming the proportions remained the same) because the numerator term $p_i(1 - p_i)$ is greater for men than women.

8.71. We test H_0: $p_1 = p_2$ versus H_a: $p_1 \neq p_2$. With $\hat{p}_1 \doteq 0.5281$, $\hat{p}_2 \doteq 0.4790$, and $\hat{p} = \frac{1024+285}{1939+595} \doteq 0.5166$, we find $SE_{D_p} \doteq 0.02342$, so $z = (\hat{p}_1 - \hat{p}_2)/SE_{D_p} \doteq 2.10$. This gives $P = 0.0360$—significant evidence (at the 5% level) that a higher proportion of non-pet owners are women.

8.72. For each confidence interval, the standard error is

$$SE_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{1102}}$$

and the margin of error is $1.96\, SE_{\hat{p}}$.

Genre	$\hat{p}$	SE	m.e.	Interval
Racing	0.74	0.01321	0.02590	0.7141 to 0.7659
Puzzle	0.72	0.01353	0.02651	0.6935 to 0.7465
Sports	0.68	0.01405	0.02754	0.6525 to 0.7075
Action	0.67	0.01416	0.02776	0.6422 to 0.6978
Adventure	0.66	0.01427	0.02797	0.6320 to 0.6880
Rhythm	0.61	0.01469	0.02880	0.5812 to 0.6388

8.73. **(a)** While there is only a 5% chance of any interval being wrong, we have six (roughly independent) chances to make that mistake. **(b)** For 99.2% confidence, use $z^* = 2.65$. (Using software, $z^* \doteq 2.6521$, or 2.6383 using the exact value of $0.05/6 = 0.008\overline{3}$). **(c)** The margin of error for each interval is $z^* SE_{\hat{p}}$, so each interval is about 1.35 times wider than in the previous exercise. (If intervals are rounded to three decimal places, as on the right, the results are the same regardless of the value of z^* used.)

Genre	Interval
Racing	0.705 to 0.775
Puzzle	0.684 to 0.756
Sports	0.643 to 0.717
Action	0.632 to 0.708
Adventure	0.622 to 0.698
Rhythm	0.571 to 0.649

8.74. **(a)** The proportion is $\hat{p} = 0.042$, so $X = (0.042)(15,000) = 630$ of the households in the sample are wireless only. **(b)** $SE_{\hat{p}} \doteq 0.00164$, so the 95% confidence interval is $0.042 \pm 0.00321 = 0.0388$ to 0.0452.

8.75. **(a)** The proportion is $\hat{p} = 0.164$, so $X = (0.164)(15,000) = 2460$ of the households in the sample are wireless only. **(b)** $SE_{\hat{p}} \doteq 0.00302$, so the 95% confidence interval is $0.164 \pm 0.00593 = 0.1581$ to 0.1699. **(c)** The estimate is 16.4%, and the interval is 15.8%

to 17.0%. **(d)** The difference in the sample proportions is $D = 0.164 - 0.042 = 0.122$. **(e)** $\text{SE}_D \doteq 0.00344$, so the margin of error is $1.96\,\text{SE}_D \doteq 0.00674$. (The confidence interval is therefore 0.1153 to 0.1287.)

8.76. **(a)** The "relative risk" is $\frac{0.164}{0.042} \doteq 3.90$. A better term might be "relative rate" (of being wireless only). **(b)** A possible summary: From 2003 to 2007, the proportion of wireless-only households increased by nearly four times. (With software, and/or the methods of Chapter 14, one could also determine a confidence interval for this ratio, as in Example 8.11.) **(c)** Both illustrate (in different ways) a change in the proportion. **(d)** Preferences will vary. Students might note that ratios are often used in news reports; for example, "the risk of complications is twice as high for people taking drug A compared to those taking drug B."

8.77. With $\hat{p}_1 = 0.43$, $\hat{p}_2 = 0.32$, and $n_1 = n_2 = 1430$, we have $\text{SE}_D \doteq 0.01799$, so the 95% confidence interval is $(0.43 - 0.32) \pm 0.03526 = 0.0747$ to 0.1453.

8.78. The pooled estimate of the proportion is $\hat{p} = 0.375$ (the average of $\hat{p}_1$ and $\hat{p}_2$, because the sample sizes were equal). Then $\text{SE}_{D_p} \doteq 0.01811$, so $z = (0.43 - 0.32)/\text{SE}_{D_p} \doteq 6.08$, for which $P < 0.0001$.

8.79. **(a)** and **(b)** The revised confidence intervals and z statistics are in the table below. **(c)** While the interval and z statistic change slightly, the conclusions are roughly the same.

Note: *Even if the second sample size were as low as 100, the two proportions would be significantly different, albeit less so ($z = 2.15$, $P = 0.0313$).*

n_2	SE_D	m.e.	Interval	$\hat{p}$	SE_{D_p}	z	P
1000	0.01972	0.03866	0.0713 to 0.1487	0.3847	0.02006	5.48	< 0.0001
2000	0.01674	0.03281	0.0772 to 0.1428	0.3659	0.01668	6.59	< 0.0001

8.80. With $\hat{p} = \frac{994}{1430} \doteq 0.6951$, we have $\text{SE}_{\hat{p}} \doteq 0.01217$, so the 95% confidence interval is $\hat{p} \pm 1.96\,\text{SE}_{\hat{p}} \doteq 0.6951 \pm 0.0239 = 0.6712$ to 0.7190.

8.81. Student answers will vary. Shown on the right is the margin of error arising for sample sizes ranging from 500 to 2300; a graphical summary is not shown, but a good choice would be a plot of margin of error versus sample size.

n	m.e.
500	0.04035
800	0.03190
1100	0.02721
1430	0.02386
1700	0.02188
2000	0.02018
2300	0.01881

8.82. With $\hat{p} = 0.58$, the standard error is $\text{SE}_{\hat{p}} = \sqrt{(0.58)(0.42)/1048} \doteq 0.01525$, so the margin of error for 90% confidence is $1.645\,\text{SE}_{\hat{p}} \doteq 0.02508$, and the interval is 0.5549 to 0.6051.

8.83. With $\hat{p}_m = 0.59$ and $\hat{p}_w = 0.56$, the standard error is $\text{SE}_D \doteq 0.03053$, the margin of error for 95% confidence is $1.96\text{SE}_D \doteq 0.05983$, and the confidence interval for $p_m - p_w$ is -0.0298 to 0.0898.

8.84. **(a)** The table below summarizes the margins of error m.e.$= 1.96\sqrt{\hat{p}(1-\hat{p})/n}$:

		$\hat{p}$	m.e.	95% confidence interval
Current downloaders ($n = 247$)	Downloading less	38%	6.05%	31.95% to 44.05%
	Use P2P networks	33.33%	5.88%	27.45% to 39.21%
	Use e-mail/IM	24%	5.33%	18.67% to 29.33%
	Use music-related sites	20%	4.99%	15.01% to 24.99%
	Use paid services	17%	4.68%	12.32% to 21.68%
All users ($n = 1371$)	Have used paid services	7%	1.35%	5.65% to 8.35%
	Currently use paid services	3%	0.90%	2.10% to 3.90%

(b) Obviously, students' renditions of the above information in a paragraph will vary.
(c) Student opinions may vary on this. Personally, I lean toward B, although I would be inclined to report two margins of error: "no more than 6%" for the current downloaders and "no more than 1.4%" for all users.

8.85. **(a)** People have different symptoms; for example, not all who wheeze consult a doctor.
(b) In the table (below), we find for "sleep" that $\hat{p}_1 = \frac{45}{282} \doteq 0.1596$ and $\hat{p}_2 = \frac{12}{164} \doteq 0.0732$, so the difference is $\hat{p}_1 - \hat{p}_2 \doteq 0.0864$. Therefore, $\text{SE}_D \doteq 0.02982$ and the margin of error for 95% confidence is 0.05844. Other computations are performed in like manner.
(c) It is reasonable to expect that the bypass proportions would be higher—that is, we expect more improvement where the pollution decreased—so we could use the alternative $p_1 > p_2$. **(d)** For "sleep," we find $\hat{p} = \frac{45+12}{282+164} \doteq 0.1278$ and $\text{SE}_{D_p} \doteq 0.03279$. Therefore, $z \doteq (0.1596 - 0.0732)/\text{SE}_{D_p} \doteq 2.64$. Other computations are similar. Only the "sleep" difference is significant. **(e)** 95% confidence intervals are shown below. **(f)** Part (b) showed improvement relative to control group, which is a better measure of the effect of the bypass, because it allows us to account for the improvement reported over time even when no change was made.

	Bypass minus congested				Bypass	
Complaint	$\hat{p}_1 - \hat{p}_2$	95% CI	z	P	$\hat{p}$	95% CI
Sleep	0.0864	0.0280 to 0.1448	2.64	0.0042	0.1596	0.1168 to 0.2023
Number	0.0307	−0.0361 to 0.0976	0.88	0.1897	0.1596	0.1168 to 0.2023
Speech	0.0182	−0.0152 to 0.0515	0.99	0.1600	0.0426	0.0190 to 0.0661
Activities	0.0137	−0.0395 to 0.0670	0.50	0.3100	0.0925	0.0586 to 0.1264
Doctor	−0.0112	−0.0796 to 0.0573	−0.32	0.6267	0.1174	0.0773 to 0.1576
Phlegm	−0.0220	−0.0711 to 0.0271	−0.92	0.8217	0.0474	0.0212 to 0.0736
Cough	−0.0323	−0.0853 to 0.0207	−1.25	0.8950	0.0575	0.0292 to 0.0857

8.86. **(a)** The number of orders completed in five days or less before the changes was $X_1 = (0.16)(200) = 32$. With $\hat{p}_1 = 0.16$, $\text{SE}_{\hat{p}} \doteq 0.02592$, and the 95% confidence interval for p_1 is 0.1092 to 0.2108. **(b)** After the changes, $X_2 = (0.9)(200) = 180$. With $\hat{p}_2 = 0.9$, $\text{SE}_{\hat{p}} \doteq 0.02121$, and the 95% confidence interval for p_2 is 0.8584 to 0.9416.
(c) $\text{SE}_D \doteq 0.03350$ and the 95% confidence interval for $p_2 - p_1$ is 0.6743 to 0.8057, or about 67.4% to 80.6%.

8.87. With $\hat{p} = 0.56$, $\text{SE}_{\hat{p}} \doteq 0.01433$, so the margin of error for 95% confidence is $1.96\text{SE}_{\hat{p}} \doteq 0.02809$.

8.88. (a) $X_1 = 121 \doteq (0.903)(134)$ die-hard fans and $X_2 = 161 \doteq (0.679)(237)$ less loyal fans watched or listened as children. **(b)** $\hat{p} = \frac{121+161}{134+237} \doteq 0.7601$ and $\text{SE}_{D_p} \doteq 0.04615$, so we find $z \doteq 4.85$ $(P < 0.0001)$—strong evidence of a difference in childhood experience. **(c)** For a 95% confidence interval, $\text{SE}_D \doteq 0.03966$ and the interval is 0.1459 to 0.3014. If students work with the rounded proportions (0.903 and 0.679), the 95% confidence interval is 0.1463 to 0.3017.

8.89. With $\hat{p}_1 = \frac{2}{3}$ and $\hat{p}_2 = 0.2$, we have $\hat{p} = \frac{134\hat{p}_1 + 237\hat{p}_2}{134+237} \doteq 0.3686$, $\text{SE}_{D_p} \doteq 0.05214$, and $z = 8.95$—very strong evidence of a difference. (If we assume that "two-thirds of the die-hard fans" and "20% of the less loyal fans" mean 89 and 47 fans, respectively, then $\hat{p} \doteq 0.3666$ and $z \doteq 8.94$; the conclusion is the same.) For a 95% confidence interval, $\text{SE}_D \doteq 0.04831$ and the interval is 0.3720 to 0.5613. (With $X_1 = 89$ and $X_2 = 47$, the interval is 0.3712 to 0.5606.)

8.90. We test H_0: $p_f = p_m$ versus H_a: $p_f \neq p_m$ for each text, where, for example, p_f is the proportion of juvenile female references. We can reject H_0 for texts 2, 3, 6, and 10. The last three texts do not stand out as different from the first seven. Texts 7 and 9 are notable as the only two with a majority of juvenile male references, while six of the ten texts had juvenile female references a majority of the time.

Text	$\hat{p}_f$	$\hat{p}_m$	$\hat{p}$	z	P
1	.4000	.2059	.2308	0.96	.3361
2	.7143	.2857	.3286	2.29	.0220
3	.4464	.2154	.3223	2.71	.0067
4	.1447	.1210	.1288	0.51	.6123
5	.6667	.2791	.3043	1.41	.1584
6	.8000	.3939	.5208	5.22	.0000
7	.9500	.9722	.9643	−0.61	.5437
8	.2778	.1818	.2157	0.80	.4259
9	.6667	.7273	.7097	−0.95	.3399
10	.7222	.2520	.3103	4.04	.0001

8.91. The proportions, z-values, and P-values are:

Text	1	2	3	4	5	6	7	8	9	10
$\hat{p}$	.8718	.9000	.5372	.6738	.9348	.6875	.6429	.6471	.7097	.8759
z	4.64	6.69	0.82	5.31	5.90	5.20	3.02	2.10	6.60	9.05
P	≈ 0	≈ 0	.4133	≈ 0	≈ 0	≈ 0	.0025	.0357	≈ 0	≈ 0

We reject H_0: $p = 0.5$ for all texts except Text 3 and (perhaps) Text 8. If we are using a "multiple comparisons" procedure such as Bonferroni (see Chapter 6), we also might fail to reject H_0 for Text 7.

The last three texts do not seem to be any different from the first seven; the gender of the author does not seem to affect the proportion.

8.92. (a) $\hat{p} = \frac{463}{975} \doteq 0.4749$, $\text{SE}_D \doteq 0.01599$, and the 95% confidence interval is 0.4435 to 0.5062. **(b)** Expressed as percents, the confidence interval is 44.35% to 50.62%. **(c)** Multiply the upper and lower limits of the confidence interval by 37,500: about 16,632 to 18,983 students.

8.94. With sample sizes of $n_w = (0.52)(1200) = 624$ women and $n_m = 576$ men, we test H_0: $p_m = p_w$ versus H_a: $p_m \neq p_w$. Assuming there were $X_m = 0.62n_m = 357.12$ men and $X_w = 0.51n_w = 318.24$ women who thought that parents put too little pressure on students, the pooled estimate is $\hat{p} \doteq 0.5628$, $\text{SE}_{D_p} \doteq 0.02866$, and the test statistic is $z \doteq 3.84$. This is strong evidence ($P < 0.0001$) that a higher proportion of men have this opinion.

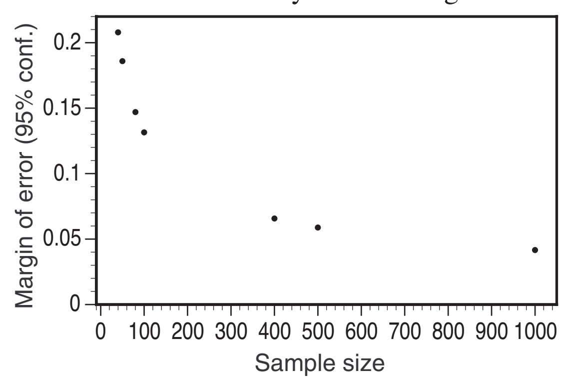
To construct a 95% confidence interval for $p_m - p_w$, we have $\text{SE}_D \doteq 0.02845$, yielding the interval 0.0542 to 0.1658.

8.95. The difference becomes more significant (i.e., the P-value decreases) as the sample size increases. For small sample sizes, the difference between $\hat{p}_1 = 0.5$ and $\hat{p}_2 = 0.4$ is not significant, but with larger sample sizes, we expect that the sample proportions should be better estimates of their respective population proportions, so $\hat{p}_1 - \hat{p}_2 = 0.1$ suggests that $p_1 \neq p_2$.

n	z	P
40	0.90	0.3681
50	1.01	0.3125
80	1.27	0.2041
100	1.42	0.1556
400	2.84	0.0045
500	3.18	0.0015
1000	4.49	0.0000

8.96. The table and graph below show the large-sample margins of error. The margin of error decreases as sample size increases, but the rate of decrease is noticeably less for large n.

n	m.e.
40	0.2079
50	0.1859
80	0.1470
100	0.1315
400	0.0657
500	0.0588
1000	0.0416



8.97. **(a)** Using either trial and error, or the formula derived in part (b), we find that at least $n = 342$ is needed. **(b)** Generally, the margin of error is $m = z^*\sqrt{\dfrac{\hat{p}_1(1-\hat{p}_1)}{n} + \dfrac{\hat{p}_2(1-\hat{p}_2)}{n}}$; with $\hat{p}_1 = \hat{p}_2 = 0.5$, this is $m = z^*\sqrt{0.5/n}$. Solving for n, we find $n = (z^*/m)^2/2$.

8.98. We must assume that we can treat the births recorded during these two times as independent SRSs. Note that the rules of thumb for the Normal approximation are not satisfied here; specifically, three birth defects are less than ten. Additionally, one might call into question the assumption of independence, because there may have been multiple births to the same set of parents included in these counts (either twins/triplets/etc., or "ordinary" siblings).

If we carry out the analysis in spite of these issues, we find $\hat{p}_1 = \frac{16}{414} \doteq 0.03865$ and $\hat{p}_2 = \frac{3}{228} \doteq 0.01316$. We might then find a 95% confidence interval: $\text{SE}_D \doteq 0.01211$, so the interval is $\hat{p}_1 - \hat{p}_2 \pm (1.96)(0.01211) = 0.00175$ to 0.04923. Note that this does not take into account the presumed direction of the difference. (This setting does meet our requirements for the plus four method, for which $\tilde{p}_1 = 0.04086$ and $\tilde{p}_2 = 0.01739$, $\text{SE}_D = 0.01298$, and the 95% confidence interval is -0.0020 to 0.0489.)

We could also perform a significance test of H_0: $p_1 = p_2$ versus H_a: $p_1 > p_2$: $\hat{p} = \frac{19}{642} \doteq 0.02960$, $\mathrm{SE}_{D_p} \doteq 0.01398$, $z \doteq 1.82$, $P = 0.0344$.

Both the large-sample interval and the significance test suggest that the two proportions are different (but not much); the plus four interval does not establish that $p_1 \neq p_2$. Also, we must recognize that the issues noted above make this conclusion questionable.

8.99. (a) $p_0 = \frac{143{,}611}{181{,}535} \doteq 0.7911$. **(b)** $\hat{p} = \frac{339}{870} \doteq 0.3897$, $\sigma_{\hat{p}} \doteq 0.0138$, and $z = (\hat{p} - p_0)/\sigma_{\hat{p}} \doteq -29.1$, so $P \doteq 0$ (regardless of whether H_a is $p > p_0$ or $p \neq p_0$). This is very strong evidence against H_0; we conclude that Mexican Americans are underrepresented on juries. **(c)** $\hat{p}_1 = \frac{339}{870} \doteq 0.3897$, while $\hat{p}_2 = \frac{143{,}611 - 339}{181{,}535 - 870} \doteq 0.7930$. Then $\hat{p} \doteq 0.7911$ (the value of p_0 from part (a)), $\mathrm{SE}_{D_p} \doteq 0.01382$, and $z \doteq -29.2$—and again, we have a tiny P-value and reject H_0.

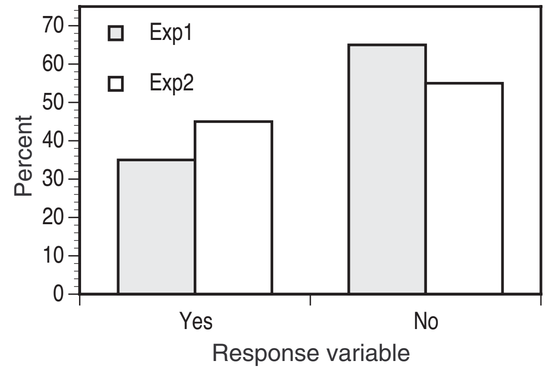
8.101. In each case, the standard error is $\sqrt{\hat{p}(1-\hat{p})/1280}$. One observation is that, while many feel that loans are a burden and wish they had borrowed less, a majority are satisfied with the benefits they receive from their education.

	$\hat{p}$	$\mathrm{SE}_{\hat{p}}$	95% confidence interval
Burdened by debt	0.555	0.01389	0.5278 to 0.5822
Would borrow less	0.544	0.01392	0.5167 to 0.5713
More hardship	0.343	0.01327	0.3170 to 0.3690
Loans worth it	0.589	0.01375	0.5620 to 0.6160
Career opportunities	0.589	0.01375	0.5620 to 0.6160
Personal growth	0.715	0.01262	0.6903 to 0.7397

Chapter 9 Solutions

9.1. **(a)** The conditional distributions are given in the table below. For example, given that Explanatory = 1, the distribution of the response variable is $\frac{70}{200} = 35\%$ "Yes" and $\frac{130}{200} = 65\%$ "No." **(b)** The graphical display might take the form of a bar graph like the one shown below, but other presentations are possible. **(c)** One notable feature is that when Explanatory = 1, "No" is more common, but "Yes" and "No" are closer to being evenly split when Explanatory = 2.

Response variable	Explanatory variable 1	2
Yes	35%	45%
No	65%	55%



9.2. **(a)** The expected count for the first cell is $\frac{(160)(200)}{400} = 80$. **(b)** This X^2 statistic has df $= (2-1)(2-1) = 1$. **(c)** Because $3.84 < X^2 < 5.02$, the P-value is between 0.025 and 0.05.

9.3. The relative risk is $\frac{0.00753}{0.00899} = 0.838$. Example 9.7 gave the 95% confidence interval $(1.02, 1.32)$, so with the ratio reversed, the interval would be approximately $(0.758, 0.980)$. For this relative risk, the statement made in Example 9.7 would be (changes underlined): "Since this interval does not include the value 1, corresponding to equal proportions in the two groups, we conclude that the lower CVD rate is statistically significant with $P < 0.05$. The low salt diet is associated with a 16% lower rate of CVD than the high salt diet."

9.4. The nine terms are shown in the table on the right. For example, the first term is

$$\frac{(69 - 51.90)^2}{51.90} \doteq 5.6341$$

These terms add up to about 14.1558; the slight difference is due to the rounding of the expected values reported in Example 9.10.

Fruit consumption	Physical activity: Low	Moderate	Vigorous
Low	5.6341	0.2230	0.3420
Medium	0.6256	0.2898	0.0153
High	6.1280	0.0091	0.8889

9.5. The table below summarizes the bounds for the P-values, and also gives the exact P-values (given by software). In each case, df $= (r-1)(c-1)$.

	X^2	Size of table	df	Crit. values (Table F)	Bounds for P	Actual P
(a)	5.32	2 by 2	1	$5.02 < X^2 < 5.41$	$0.02 < P < 0.025$	0.0211
(b)	2.7	2 by 2	1	$2.07 < X^2 < 2.71$	$0.10 < P < 0.15$	0.1003
(c)	25.2	4 by 5	12	$24.05 < X^2 < 26.22$	$0.01 < P < 0.02$	0.0139
(d)	25.2	5 by 4	12	$24.05 < X^2 < 26.22$	$0.01 < P < 0.02$	0.0139

9.6. The Minitab output shown on the right gives $X^2 \doteq 54.307$, df $= 1$, and $P < 0.0005$, indicating significant evidence of an association.

Minitab output

```
          Men    Women    Total
 Yes     1392     1748     3140
      1215.19  1924.81

  No     3956     6723    10679
      4132.81  6546.19

Total    5348     8471    13819

ChiSq = 25.726 + 16.241 +
         7.564 +  4.776 = 54.307
df = 1, p = 0.000
```

9.7. The expected counts were rounded to the nearest hundredth.

9.8. The table below lists the observed counts, the population proportions, the expected counts, and the chi-square contributions (for the next exercise). Each expected count is the product of the proportion and the sample size 1567; for example, $(0.172)(1567) = 269.524$ for California.

State	AZ	CA	HI	IN	NV	OH
Observed count	167	257	257	297	107	482
Proportion	0.105	0.172	0.164	0.188	0.070	0.301
Expected count	164.535	269.524	256.988	294.596	109.690	471.667
Chi-square contribution	0.0369	0.5820	0.0000	0.0196	0.0660	0.2264

9.9. The expected counts are in the table above, rounded to four decimal places as in Example 9.15; for example, for California, we have

$$\frac{(257 - 269.524)^2}{269.524} \doteq 0.5820$$

The six values add up to 0.93 (rounded to two decimal places).

9.10. The chi-square goodness of fit statistic is $X^2 \doteq 15.2$ with df $= 5$, for which $0.005 < P < 0.01$ (software gives 0.0096). The details of the computation are given in the table below; note that there were 475 M&M's in the bag.

	Expected frequency	Expected count	Observed count	$O - E$	$\frac{(O-E)^2}{E}$
Brown	0.13	61.75	61	−0.75	0.0091
Yellow	0.14	66.5	59	−7.5	0.8459
Red	0.13	61.75	49	−12.75	2.6326
Orange	0.20	95	77	−18	3.4105
Blue	0.24	114	141	27	6.3947
Green	0.16	76	88	12	1.8947
			475		15.1876

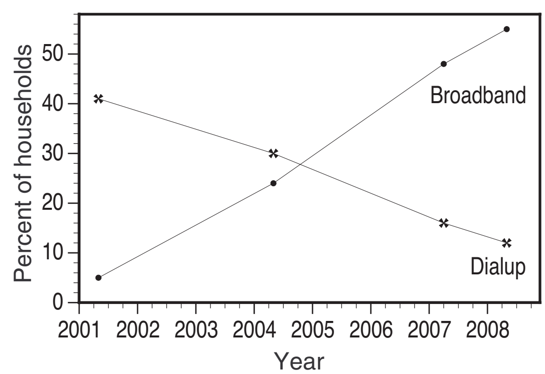
9.11. **(a)** The two-way table is on the right; for example, for April 2001, $(0.05)(2250) = 112.5$ and $(0.95)(2250) = 2137.5$. **(b)** Under the null hypothesis that the proportions have not changed, the expected counts are $(0.33)(2250) = 742.5$ (across the top row) and $(0.67)(2250) = 1507.5$ (across the bottom row), because the average of the four broadband percents is $\frac{5\%+24\%+48\%+55\%}{4} = 33\%$. (We take the unweighted average because we have assumed that the sample sizes were equal.) The test statistic is $X^2 \doteq 1601.8$ with df = 3, for which $P < 0.0001$. Not surprisingly, we reject H_0. **(c)** The average of the last two broadband percents is $\frac{48\%+55\%}{2} = 51.5\%$, so if the proportions are equal, the expected counts are $(0.515)(2250) = 1158.75$ (top row) and $(0.485)(2250) = 1091.25$ (bottom row). The test statistic is $X^2 \doteq 22.07$ with df = 1, for which $P < 0.0001$.

	Date of Survey			
Broadband?	April 2001	April 2004	March 2007	April 2008
Yes	112.5	540	1080	1237.5
No	2137.5	1710	1170	1012.5

Note: *This test is equivalent to testing H_0: $p_1 = p_2$ versus H_a: $p_1 \neq p_2$ using the methods of Chapter 8. We find pooled estimate $\hat{p} = 0.515$, $SE_{D_p} \doteq 0.01490$, and $z = (0.48 - 0.55)/SE_{D_p} \doteq -4.70$. (Note that $z^2 = X^2$.)*

9.12. **(a)** The two-way table is on the right; for example, for April 2001, $(0.41)(2250) = 922.5$ and $(0.59)(2250) = 1327.5$. **(b)** Under the null hypothesis that the proportions have not changed, the expected counts are $(0.2475)(2250) = 556.875$ (across the top row) and $(0.7525)(2250) = 1693.125$ (across the bottom row), because the average of the four dialup percents is $\frac{41\%+30\%+16\%+12\%}{4} = 24.75\%$. The test statistic is $X^2 \doteq 641.2$ with df = 3, for which $P < 0.0001$. Again, we reject H_0. **(c)** The average of the last two dialup percents is $\frac{16\%+12\%}{2} = 14\%$, so if the proportions are equal, the expected counts are $(0.14)(2250) = 315$ (top row) and $(0.86)(2250) = 1935$ (bottom row). The test statistic is $X^2 \doteq 14.95$ with df = 1, for which $P < 0.0001$. **(d)** The data shows that the rise of broadband access has been accompanied by a decline in dialup access.

	Date of Survey			
Dialup?	April 2001	April 2004	March 2007	April 2008
Yes	922.5	675	360	270
No	1327.5	1575	1890	1980



Note: *As in the previous exercise, the test in part (c) is equivalent to testing H_0: $p_1 = p_2$ versus H_a: $p_1 \neq p_2$, for which the pooled estimate is $\hat{p} = 0.14$, $SE_{D_p} \doteq 0.01035$, and $z = (0.16 - 0.12)/SE_{D_p} \doteq 3.87$. Again, note that $z^2 = X^2$.)*

9.13. Students may experiment with a variety of scenarios, but they should find that regardless of the what they try, the conclusion is the same.

9.14. **(a)** Student approaches to estimating the dialup counts will vary. The bottom row of the table on the right shows a reasonable set of estimates, found by fitting a regression line to the counts in the solution to Exercise 9.13.

	Date of Survey			
Switch?	October 2002	February 2004	December 2005	May 2008
Yes	301.10	266.29	191.71	94.32
No	491.28	399.43	299.86	167.68
Total	792.38	665.72	491.57	262.00

(Even students who use a similar approach might get slightly different answers depending on how they represent the survey dates as x values.) **(b)** For example, for October 2002, $(0.38)(792.38) = 301.10$ and $(0.62)(792.38) = 491.28$. **(c)** For the data shown, the test statistic is $X^2 \doteq 1.45$ (df $= 3$, $P = 0.6934$). Student results will vary, but unless their dialup count estimates are drastically different, they should not reject H_0; that is, there is not enough evidence to conclude that the proportion of dialup users intending to switch to broadband has changed. **(d)** Answers will vary depending on the approach used, but should be close to 45%. One explanation is that the number of (surveyed) dialup users who were not interested in switching dropped from about 300 to 168 from December 2005 to May 2008—a 44% reduction. Alternatively, in that time period, the number of dialup users dropped by 47%, from about 492 to 262. In order for the percent not planning to switch to remain at 60%, that group must decrease by a similar amount.

9.15. **(a)** The 3×2 table is on the right. **(b)** The percents of disallowed small, medium, and large claims are (respectively) $\frac{6}{57} \doteq 10.5\%$, $\frac{5}{17} \doteq 29.4\%$, and $\frac{1}{5} = 20\%$. **(c)** In the 3×2 table, the expected count for large/not allowed is too small $(\frac{5 \cdot 12}{79} \doteq 0.76)$. **(d)** The null hypothesis is "There is no relationship between claim size and whether a claim is allowed."

	Allowed?		
Stratum	Yes	No	Total
Small	51	6	57
Medium	12	5	17
Large	4	1	5
Total	67	12	79

(e) As a 2×2 table (with the second row 16 "yes" and 6 "no"), we find $X^2 = 3.456$, df $= 1$, $P = 0.063$. The evidence is not quite strong enough to reject H_0.

9.16. **(a)** In the table below, the estimated numbers of disallowed claims in the populations are found by multiplying the sample proportion by the population size; for example, $\frac{6}{57} \cdot 3342 \doteq 351.8$ claims. **(b)** For each stratum, let $\hat{p}$ be the sample proportion, n be the sample size, and N be the population size. The standard error for the sample is $\mathrm{SE}_{\hat{p}} = \sqrt{\hat{p}(1-\hat{p})/n}$, and the standard error for the population estimate is $N\,\mathrm{SE}_{\hat{p}}$. The margins of error depends on the desired confidence level; for 95% confidence, we should double the population standard errors.

	Sample		Population		Standard error	
Stratum	Not allowed	Total	Not allowed	Total	Sample	Population
Small	6	57	351.8	3342	0.0406	135.8485
Medium	5	17	72.4	246	0.1105	27.1855
Large	1	5	11.6	58	0.1789	10.3754

9.17. The table on the right shows the given information translated into a 3×2 table. For example, in Year 1, about $(0.423)(2408) = 1018.584$ students received DFW grades, and the rest—$(0.577)(2408) = 1389.416$ students—passed. To test H_0: the DFW rate has not changed, we have $X^2 \doteq 307.8$, df = 2, $P < 0.0001$—very strong evidence of a change.

Year	DFW	Pass
1	1018.584	1389.416
2	578.925	1746.075
3	423.074	1702.926

9.18. (a) The table of approximate counts is on the right. Because the reported percents were rounded to the nearest whole percent, the total sample size is *not* 719. **(b)** With the counts as in the table, $X^2 \doteq 15.75$, df = 3, and $P \doteq 0.0013$. If students round the counts, or attempt to adjust the numbers in the first column so the numbers add up to 719, the value of X^2 will change slightly, but the P-value remains small, and the conclusion is the same. **(c)** We have strong enough evidence to conclude that there is an association between class attendance and DFW rates. **(d)** Association is not proof of causation. However, by comparing the observed counts with the expected counts, we can see that the data are consistent with that scenario; for example, among students with the highest attendance rates, more passed than expected (355.74 observed, 336.33 expected), and fewer failed (91.8 observed, 111.2 expected).

Attendance	ABC	DFW	Total
Less than 50%	10.78	9	19.78
51% to 74%	43.12	25.2	68.32
75% to 94%	134.75	54	188.75
95% or more	355.74	91.8	447.54
Total	544.39	180	724.39

9.19. (a) The approximate counts are shown on the right; for example, among those students in trades, $(0.34)(942) = 320.28$ enrolled right after high school, and $(0.66)(942) = 621.72$ enrolled later. **(b)** In addition to a chi-square test in part (c), students might note other things, such as: Overall, 39.4% of these students enrolled right after high school. Health is the most popular field, with about 38% of these students. **(c)** We have strong enough evidence to conclude that there is an association between field of study and when students enter college; the test statistic is $X^2 = 275.9$ (with unrounded counts) or 276.1 (with rounded counts), with df = 5, for which P is very small. A graphical summary is not shown; a bar chart would be appropriate.

	Time of entry		
Field of study	Right after high school	Later	Total
Trades	320.28	621.72	942
Design	274.48	309.52	584
Health	2034	3051	5085
Media/IT	975.88	2172.12	3148
Service	486	864	1350
Other	1172.60	1082.40	2255
Total	5263.24	8100.76	13,364

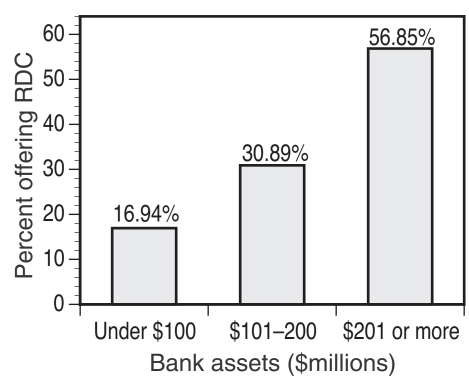
9.20. **(a)** The approximate counts are shown on the right; for example, among those students in trades, $(0.45)(942) = 423.9$ took government loans and $(0.55)(942) = 518.1$ did not. **(b)** We have strong enough evidence to conclude that there is an association between field of study and taking government loans; the test statistic is $X^2 = 97.44$ (with unrounded counts) or 97.55 (with rounded counts), with df = 5, for which P is very small. **(c)** Overall, 53.3% of these students took government loans; students in trades and "other" fields of study were slightly less likely, and those in the service field were slightly more likely. A bar graph would be a good choice for a graphical summary.

	Government loans		
Field	Yes	No	Total
Trades	423.9	518.1	942
Design	317.47	281.53	599
Health	2878.7	2355.3	5234
Media/IT	1780.9	1457.1	3238
Service	826.8	551.2	1378
Other	1081	1219	2300
Total	7308.77	6382.23	13,691

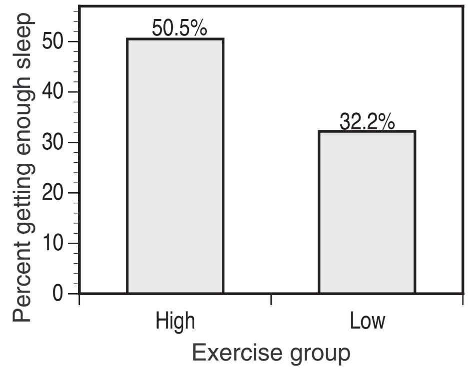
9.21. **(a)** The approximate counts are shown on the right; for example, among those students in trades, $(0.2)(942) = 188.4$ relied on parents, family, or spouse, and $(0.8)(942) = 753.6$ did not. **(b)** We have strong enough evidence to conclude that there is an association between field of study and getting money from parents, family, or spouse; the test statistic is $X^2 = 544.0$ (with unrounded counts) or 544.8 (with rounded counts), with df $= 5$, for which P is very small. **(c)** Overall, 25.4% of these students relied on family support; students in media/IT and service fields were slightly less likely, and those in the design and"other" fields were slightly more likely. A bar graph would be a good choice for a graphical summary.

	Parents, family, spouse		
Field	Yes	No	Total
Trades	188.4	753.6	942
Design	221.63	377.37	599
Health	1360.84	3873.16	5234
Media/IT	518.08	2719.92	3238
Service	248.04	1129.96	1378
Other	943	1357	2300
Total	3479.99	10211.01	13,691

9.22. **(a)** For example, $\frac{63}{63+309} \doteq 16.94\%$ of the smallest banks over RDC. The bar graph on the right is one possible graphical summary. **(b)** To test H_0: no association between bank size and offering RDC, we have $X^2 \doteq 96.3$ with df $= 2$, for which P is tiny. We have very strong evidence of an association.



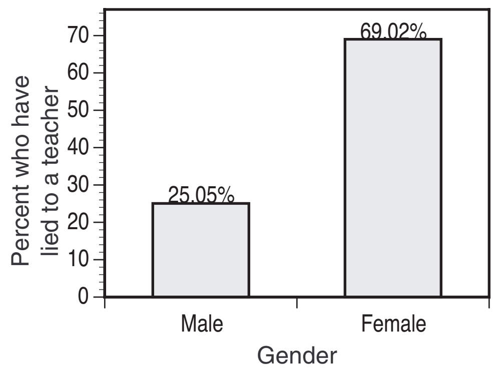
9.23. **(a)** Of the high exercisers, $\frac{151}{151+148} \doteq 50.5\%$ get enough sleep, and the rest (49.5%) do not. **(b)** Of the low exercisers, $\frac{115}{115+242} \doteq 32.2\%$ get enough sleep, and the rest (67.8%) do not. **(c)** Those who exercise more than the median are more likely to get enough sleep. **(d)** To test H_0: exercise and sleep are not associated, we have $X^2 \doteq 22.58$ with df $= 1$, for which P is very small. We have very strong evidence of an association.



9.24. **(a)** The marginal totals are given in the table on the right. **(b)** The most appropriate description is the conditional distribution by gender (the explanatory variable): 25.05% of males, and 69.02% of females, admitted to lying. **(c)** Females are much more likely to have lied (or at least, to admit to lying). **(d)** Not surprisingly, this is highly significant: $X^2 \doteq 5352$, df $= 1$, P is tiny. This test statistic is too extreme to bother creating a P-value sketch.

Note: *To get an idea of how extreme this test statistic value is: Observing* $X^2 = 5352$ *from a* $\chi^2(1)$ *distribution is equivalent to* $z = \sqrt{5352} \doteq 73$ *from the standard Normal distribution.*

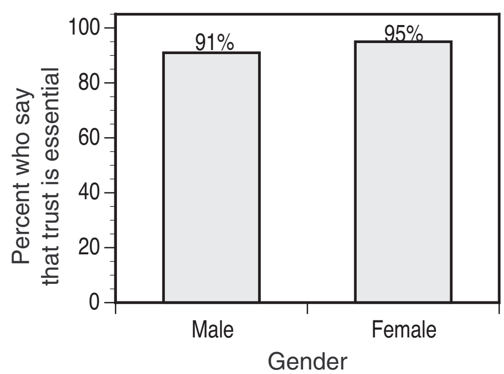
Lied?	Male	Female	Total
Yes	3,228	10,295	13,523
No	9,659	4,620	14,279
Total	12,887	14,915	27,802



9.25. **(a)** The marginal totals are given in the table on the right. **(b)** The most appropriate description is the conditional distribution by gender (the explanatory variable): 91% of males, and 95% of females, agreed that trust and honesty are essential. **(c)** Females are slightly more likely to view trust and honesty as essential. **(d)** While the percents in the conditional distribution are similar, the large sample sizes make this highly significant: $X^2 \doteq 175.0$, df $= 1$, P is tiny. Once again, a P-value sketch is not shown.

Note: $X^2 = 175$ *coming from a* $\chi^2(1)$ *distribution is equivalent to* $z = \sqrt{175} \doteq 13$ *coming from the standard Normal distribution.*

Lied?	Male	Female	Total
Yes	11,724	14,169	25,893
No	1,163	746	1,909
Total	12,887	14,915	27,802



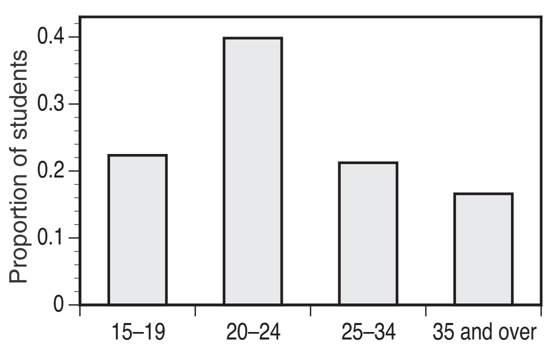
9.26. The main problem is that this is not a two-way table. Specifically, each of the 119 students might fall into several categories: They could appear on more than one row if they saw more than one of the movies and might even appear more than once on a given row (for example, if they have both bedtime and waking symptoms arising from the same movie).

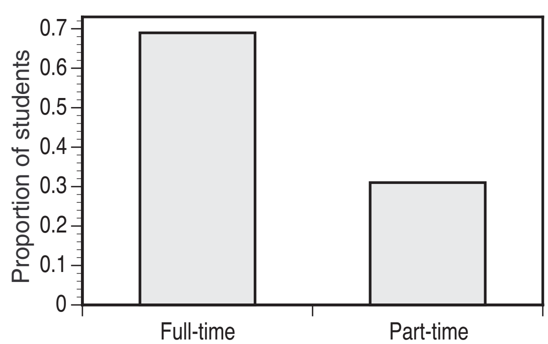
Another potential problem is that this is a table of *percents* rather than *counts*. However, because we were given the value of n for each movie title, we could use that information to determine the counts for each category; for example, it appears that 20 of the 29 students who watched *Poltergeist* had short-term bedtime problems because $\frac{20}{29} \doteq 68.96\%$ (perhaps the reported value of 68% was rounded incorrectly). If we determine all of these counts in this way (and note several more apparent rounding errors in the process), those counts add up to 200, so we see that students really were counted more than once.

If the values of n had not been given for each movie, then we could not do a chi-squared analysis even if this were a two-way table.

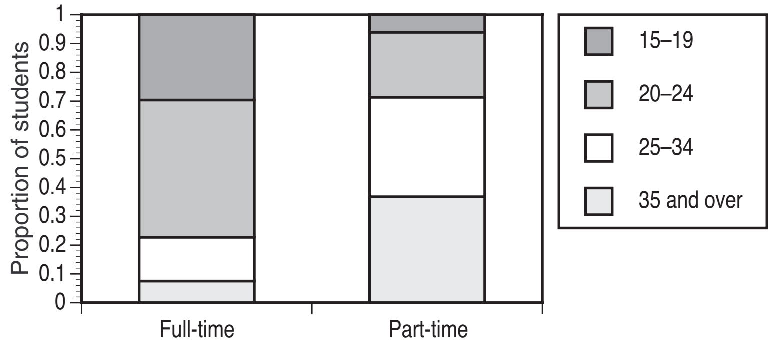
9.27. (a) The joint distribution is found by dividing each number in the table by 17,380 (the total of all the numbers). These proportions are given in italics on the right. For example, $\frac{3553}{17380} \doteq 0.2044$, meaning that about 20.4% of all college students are full-time and aged 15 to 19. **(b)** The marginal distribution of age is found by dividing the *row* totals by 17,380; they are in the right margin of the table (above, right) and the graph on the left below. For example, $\frac{3882}{17380} \doteq 0.2234$, meaning that about 22.3% of all college students are aged 15 to 19. **(c)** The marginal distribution of status is found by dividing the *column* totals by 17,380; they are in the bottom margin of the table (above, right) and the graph on the right below. For example, $\frac{11989}{17380} \doteq 0.6898$, meaning that about 69% of all college students are full-time. **(d)** The conditional distributions are given in the table on the following page. For each status category, the conditional distribution of age is found by dividing the counts in that column by that column total. For example, $\frac{3553}{11989} \doteq 0.2964$, $\frac{5710}{11989} \doteq 0.4763$, etc., meaning that of all full-time college students, about 29.64% are aged 15 to 19, 47.63% are 20 to 24, and so on. Note that each set of four numbers should add to 1 (except for rounding error). Graphical presentations may vary; one possibility is shown on the following page. **(e)** We see that full-time students are dominated by younger ages, while part-time students are more likely to be older.

	FT	PT	
15–19	3553 *0.2044*	329 *0.0189*	3882 0.2234
20–24	5710 *0.3285*	1215 *0.0699*	6925 0.3984
25–34	1825 *0.1050*	1864 *0.1072*	3689 0.2123
35+	901 *0.0518*	1983 *0.1141*	2884 0.1659
	11989 0.6898	5391 0.3102	17380

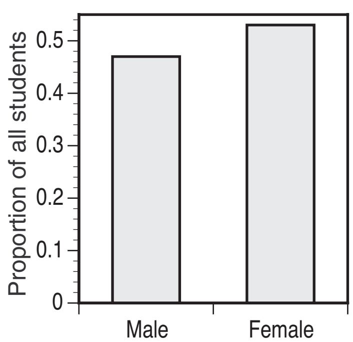


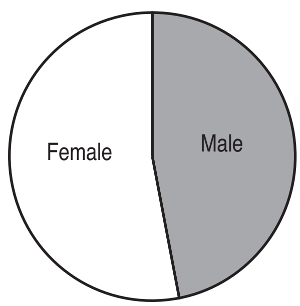


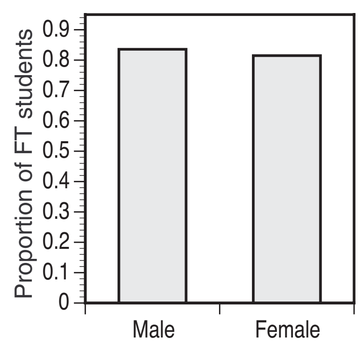
	Full-time	Part-time
15–19	0.2964	0.0610
20–24	0.4763	0.2254
25–34	0.1522	0.3458
35+	0.0752	0.3678



9.28. **(a)** Of all students aged 20 to 24 years, $\frac{3254}{6925} \doteq 46.99\%$ are men and the rest ($\frac{3671}{6925} \doteq 53.01\%$) are women. Shown below are two possible graphical displays. In the bar graph on the left, the bars represent the proportion of all students (in this age range) in each gender. Alternatively, because the two percents represent parts of a single whole, we can display the distribution as a pie chart like that in the middle. **(b)** Among male students, $\frac{2719}{3254} \doteq 83.56\%$ are full-time and the rest ($\frac{535}{3254} \doteq 16.44\%$) are part-time. Among female students, those numbers are $\frac{2991}{3671} \doteq 81.48\%$ and $\frac{680}{3671} \doteq 18.52\%$. Men in this age range are (very slightly) more likely to be full-time students. The bar graph below on the right shows the proportions of full-time students side by side; note that a pie graph would *not* be appropriate for this display because the two proportions represent parts of two different wholes. **(c)** For the full-time row, the expected counts are $\frac{(5710)(3254)}{6925} \doteq 2683.08$ and $\frac{(5710)(3671)}{6925} \doteq 3026.92$. **(d)** Using df $= 1$, we see that $X^2 = 5.17$ falls between 5.02 and 5.41, so $0.02 < P < 0.025$ (software gives 0.023). This is significant evidence (at the 5% level) that there is a difference in the conditional distributions.







9.29. **(a)** The percent who have lasting waking symptoms is the total of the first column divided by the grand total: $\frac{69}{119} \doteq 57.98\%$. **(b)** The percent who have both waking and bedtime symptoms is the count in the upper left divided by the grand total: $\frac{36}{119} \doteq 30.25\%$. **(c)** To test H_0: There is no relationship between waking and bedtime symptoms versus H_a: There is a relationship, we find $X^2 \doteq 2.275$ (df $= 1$) and $P \doteq 0.132$. We do not have enough evidence to conclude that there is a relationship.

Minitab output

```
         WakeYes   WakeNo    Total
BedYes        36       33       69
           40.01    28.99

 BedNo        33       17       50
           28.99    21.01

 Total        69       50      119

ChiSq =  0.402 +  0.554 +
         0.554 +  0.765 = 2.275
df = 1, p = 0.132
```

9.30. The table below gives df $= (r-1)(c-1)$, bounds for P, and software P-values.

	X^2	Size of table	df	Crit. values (Table F)	Bounds for P	Software P
(a)	1.25	2 by 2	1	$X^2 < 1.32$	$P > 0.25$	0.2636
(b)	18.34	4 by 4	9	$16.92 < X^2 < 19.02$	$0.025 < P < 0.05$	0.0314
(c)	24.21	2 by 8	7	$22.04 < X^2 < 24.32$	$0.001 < P < 0.0025$	0.0010
(d)	12.17	5 by 3	8	$12.03 < X^2 < 13.36$	$0.10 < P < 0.15$	0.1438

9.31. Two examples are shown on the right. In general, choose a to be any number from 0 to 50, and then all the other entries can be determined.

30	20
70	80

10	40
90	60

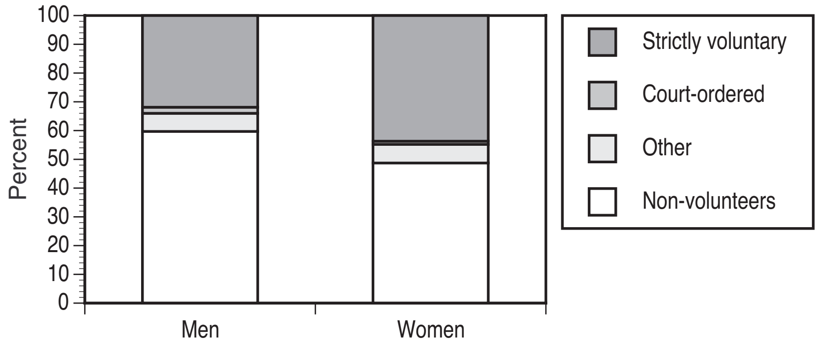
Note: *This is why we say that such a table has "one degree of freedom": We can make one (nearly) arbitrary choice for the first number, and then have no more decisions to make.*

9.32. To construct such a table, we can start by choosing values for the row and column sums r_1, r_2, r_3, c_1, c_2, as well as the grand total N. Note that the $N = r_1 + r_2 + r_3 = c_1 + c_2$, so we only have four choices to make. Then find each count a, b, c, d, e, f by taking the corresponding *row* total, times the corresponding *column* total, divided by the *grand* total. For example, $a = r_1 \times c_1/N$ and $d = r_2 \times c_2/N$. Of course, these counts should be whole numbers, so it may be necessary to make adjustments in the row and column totals to meet this requirement.

a	b	r_1
c	d	r_2
e	f	r_3
c_1	c_2	N

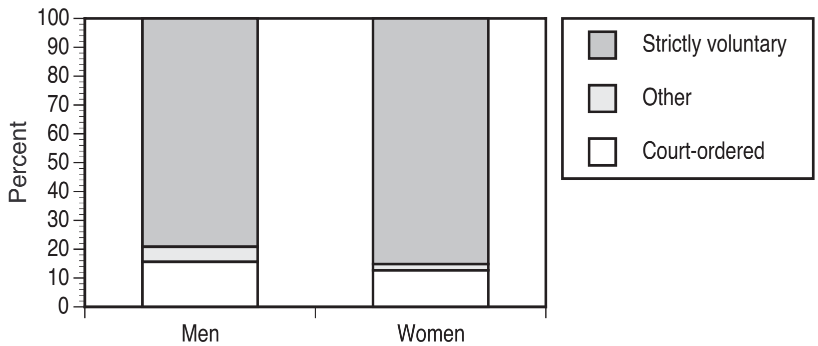
The simplest such table would have all six counts a, b, c, d, e, f equal to one another (which would arise if we start with $r_1 = r_2 = r_3$ and $c_1 = c_2$).

9.33. **(a)** Different graphical presentations are possible; one is shown on the following page. More women perform volunteer work; the notably higher percent of women who are "strictly voluntary" participants accounts for the difference. (The "court-ordered" and "other" percents are similar for men and women.) **(b)** Either by adding the three "participant" categories or by subtracting from 100% the non-participant percentage, we find that 40.3% of men and 51.3% of women are participants. The relative risk of being a volunteer is therefore $\frac{51.3\%}{40.3\%} \doteq 1.27$.



9.34. Table shown on the right; for example, $\frac{31.9\%}{40.3\%} \doteq 79.16\%$. The percents in each row sum to 100%, with no rounding error for up to four places after the decimal. Both this graph and the graph in the previous exercise show that women are more likely to volunteer, but in this view we cannot see the difference in the rate of non-participation.

Gender	Strictly voluntary	Court-ordered	Other
Men	79.16%	5.21%	15.63%
Women	85.19%	2.14%	12.67%

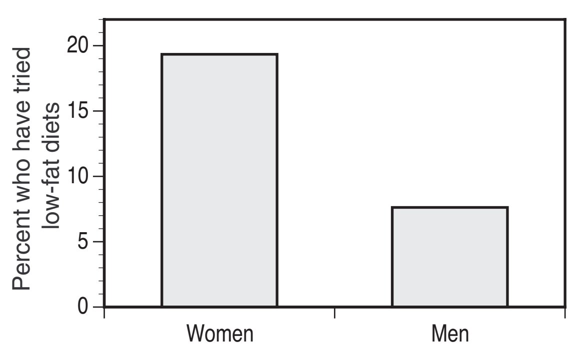


9.35. (a) The missing entries (shown shaded on the right) are found by subtracting the number who have tried low-fat diets from the given totals. **(b)** Viewing gender as explanatory, compute the conditional distributions of low-fat diet for each gender: $\frac{35}{181} \doteq 19.34\%$ of women and $\frac{8}{105} \doteq 7.62\%$ of men have tried low-fat diets. **(c)** The test statistic is $X^2 = 7.143$ (df $= 1$), for which $P = 0.008$. We have strong evidence of an association; specifically, women are more likely to try low-fat diets.

	Gender	
Low-fat diet?	Women	Men
Yes	35	8
No	146	97
Total	181	105

Minitab output

```
         Women      Men    Total
  Yes       35        8       43
         27.21    15.79

   No      146       97      243
        153.79    89.21

Total      181      105      286

ChiSq =  2.228 +  3.841 +
         0.394 +  0.680 = 7.143
df = 1, p = 0.008
```

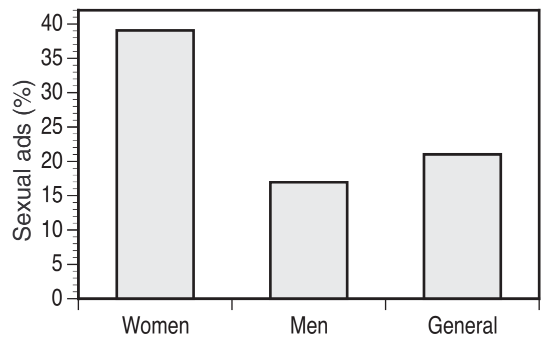


9.36. **(a)** The best numerical summary would note that we view target audience ("magazine readership") as explanatory, so we should compute the conditional distribution of model dress for each audience. This table and graph are shown below. **(b)** Minitab output is shown on the right: $X^2 \doteq 80.9$, df $= 2$, and P is very small. We have very strong evidence that target audience affects model dress. **(c)** The sample is not an SRS: A set of magazines were chosen, and then all ads in three issues of those magazines were examined. It is not clear how this sampling approach might invalidate our conclusions, but it does make them suspect.

Minitab output

```
          Women      Men     Genl    Total
    1       351      514      248     1113
         424.84   456.56   231.60

    2       225      105       66      396
         151.16   162.44    82.40

Total       576      619      314     1509

ChiSq = 12.835 +  7.227 +  1.162 +
        36.074 + 20.312 +  3.265 = 80.874
df = 2, p = 0.000
```

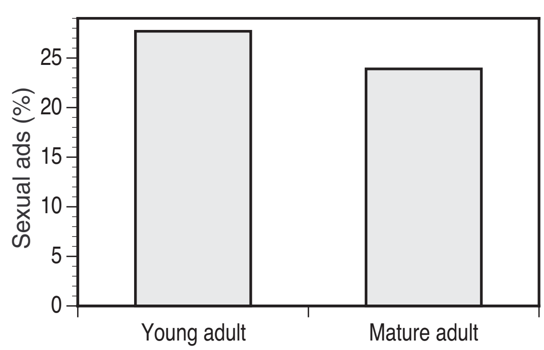
	Magazine readership		
Model dress	Women	Men	General
Not sexual	60.94%	83.04%	78.98%
Sexual	39.06%	16.96%	21.02%



9.37. **(a)** As the conditional distribution of model dress for each age group has been given to us, it only remains to display this distribution graphically. One such presentation is shown below. **(b)** In order to perform the significance test, we must first recover the counts from the percents. For example, there were $(0.723)(1006) \doteq 727$ non-sexual ads in young adult magazines. The remainder of these counts can be seen in the Minitab output below, where we see $X^2 \doteq 2.59$, df $= 1$, and $P \doteq 0.108$—not enough evidence to conclude that age group affects model dress.

Minitab output

```
          Young   Mature    Total
    1       727      383     1110
         740.00   370.00

    2       279      120      399
         266.00   133.00

Total      1006      503     1509

ChiSq =  0.228 +  0.457 +
         0.635 +  1.271 = 2.591
df = 1, p = 0.108
```



9.38. **(a)** Subtract the "agreed" counts from the sample sizes to get the "disagreed" counts. The table is in the Minitab output on the right. (The output has been slightly altered to have more descriptive row and column headings.) We find $X^2 \doteq 2.67$, df $= 1$, and $P = 0.103$, so we cannot conclude that students and non-students differ in the response to this question. **(b)** For testing H_0: $p_1 = p_2$ versus H_a: $p_1 \neq p_2$, we have $\hat{p}_1 \doteq 0.3607$, $\hat{p}_2 \doteq 0.5085$, $\hat{p} \doteq 0.4333$, $\text{SE}_{D_p} \doteq 0.09048$, and $z = -1.63$. Up to rounding, $z^2 = X^2$ and the P-values are the same. **(c)** The statistical tests in (a) and (b) assume that we have two SRSs, which we clearly do not have here. Furthermore, the two groups differed in geography (northeast/West Coast) in addition to student/non-student classification. These issues mean we should not place too much confidence in the conclusions of our significance test—or, at least, we should not generalize our conclusions too far beyond the populations "upper level northeastern college students taking a course in Internet marketing" and "West Coast residents willing to participate in commercial focus groups."

Minitab output

```
      Students   Non-st    Total
 Agr        22       30       52
         26.43    25.57

 Dis        39       29       68
         34.57    33.43

Total       61       59      120

ChiSq =  0.744 +  0.769 +
         0.569 +  0.588 = 2.669
df = 1, p = 0.103
```

9.39. **(a)** First we must find the counts in each cell of the two-way table. For example, there were about $(0.172)(5619) \doteq 966$ Division I athletes who admitted to wagering. These counts are shown in the Minitab output on the right, where we see that $X^2 \doteq 76.7$, df $= 2$, and $P < 0.0001$. There is very strong evidence that the percent of athletes who admit to wagering differs by division. **(b)** Even with much smaller numbers of students (say, 1000 from each division), P is still very small. Presumably the estimated numbers are reliable enough that we would not expect the true counts to be less than 1000, so we need not be concerned about the fact that we had to estimate the sample sizes. **(c)** If the reported proportions are wrong, then our conclusions may be suspect—especially if it is the case that athletes in some division were more likely to say they had not wagered when they had. **(d)** It is difficult to predict exactly how this might affect the results: Lack of independence could cause the estimated percents to be too large, or too small, if our sample included several athletes from teams which have (or do not have) a "gambling culture."

Minitab output

```
            Div1      Div2      Div3     Total
   1         966       621       998      2585
         1146.87    603.54    834.59

   2        4653      2336      3091     10080
         4472.13   2353.46   3254.41

Total       5619      2957      4089     12665

ChiSq = 28.525 +  0.505 + 31.996 +
         7.315 +  0.130 +  8.205 = 76.675
df = 2, p = 0.000
```

9.40. In Exercise 9.15, we are comparing three populations (model 1): small, medium, and large claims. In Exercise 9.23, we test for independence (model 2) between amount of sleep and level of exercise. In Exercise 9.24, we test for independence between gender and lying to teachers. In Exercise 9.39, one could argue for either answer. If we chose three separate random samples from each division, then we are comparing three populations (model 1). If a single random sample of student athletes was chosen, and then we classified each student by division and by gambling response, this is a test for independence (model 2).

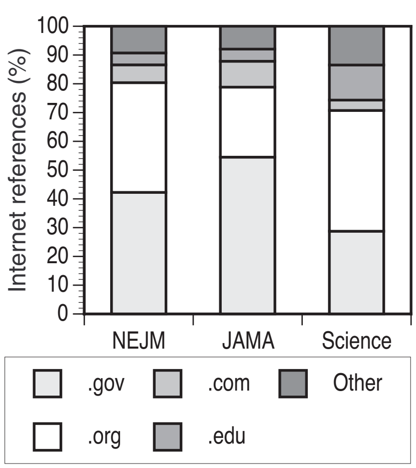
Note: *For some of these problems, either answer may be acceptable, provided a reasonable explanation is given. The distinctions between the models can be quite difficult to make since the difference between several populations might, in fact, involve classification by a categorical variable. In many ways, it comes down to how the data were collected. For example, in Exercise 9.15, we were told that the data came from a stratified random sample—which means that the three groups were treated as separate populations. Of course, the difficulty is that the method of collecting data may not always be apparent, in which case we have to make an educated guess. One question we can ask to educate our guess is whether we have data that can be used to estimate the (population) marginal distributions.*

9.41. The Minitab output on the right shows both the two-way table (column and row headings have been changed to be more descriptive) and the results for the significance test: $X^2 \doteq 12.0$, df $= 1$, and $P = 0.001$, so we conclude that gender and flower choice are related. The count of 0 does not invalidate the test: Our smallest expected count is 6, while the text says that "for 2×2 tables, we require that all four expected cell counts be 5 or more."

Minitab output

```
         Female     Male    Total
bihai        20        0       20
          14.00     6.00

   no        29       21       50
          35.00    15.00

Total        49       21       70

ChiSq =   2.571 +  6.000 +
          1.029 +  2.400 = 12.000
df = 1, p = 0.001
```

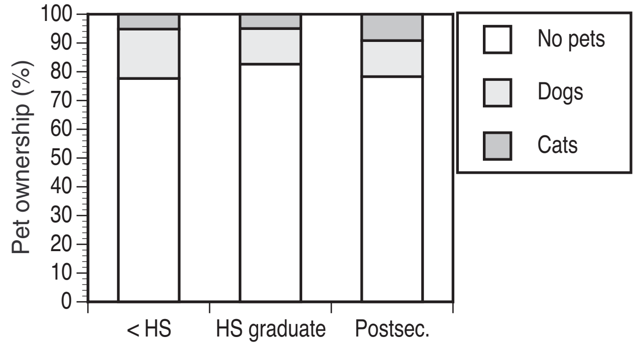
9.42. The graph below depicts the conditional distribution of domain type for each journal; for example, in *NEJM*, $\frac{41}{97} \doteq 42.27\%$ of Internet references were to .gov domains, $\frac{37}{97} \doteq 38.14\%$ were to .org domains, and so on. The Minitab output shows the expected counts, which tell a story similar to the bar graph, and show that the relationship between journal and domain type is significant ($X^2 \doteq 56.12$, df $= 8$, $P < 0.0005$).



Minitab output

```
        NEJM     JAMA  Science   Total
.gov      41      103      111     255
       36.81    71.72   146.47

.org      37       46      162     245
       35.36    68.91   140.73

.com       6       17       14      37
        5.34    10.41    21.25

.edu       4        8       47      59
        8.52    16.59    33.89

other      9       15       52      76
       10.97    21.37    43.65

Total     97      189      386     672

ChiSq = 0.477 + 13.644 +  8.591 +
        0.076 +  7.615 +  3.215 +
        0.081 +  4.178 +  2.475 +
        2.395 +  4.451 +  5.072 +
        0.354 +  1.901 +  1.595 = 56.12
df = 8, p = 0.000
```

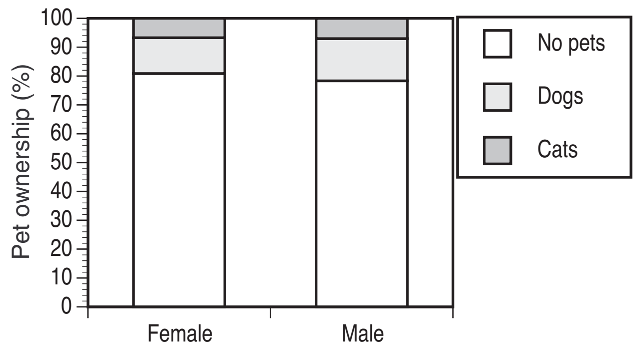
9.43. The graph on the right depicts the conditional distribution of pet ownership for each education level; for example, among those who did not finish high school, $\frac{421}{542} \doteq 77.68\%$ owned no pets, $\frac{93}{542} \doteq 17.16\%$ owned dogs, and $\frac{28}{542} \doteq 5.17\%$ (the rest) owned cats. (One could instead compute column percents—the conditional distribution of education for each pet-ownership group—but education level makes more sense as the explanatory variable here.) The (slightly altered) Minitab output shows that the relationship between education level and pet ownership is significant ($X^2 \doteq 23.15$, df $= 4$, $P < 0.0005$). Specifically, dog owners have less education, and cat owners more, than we would expect if there were no relationship between pet ownership and educational level.



Minitab output

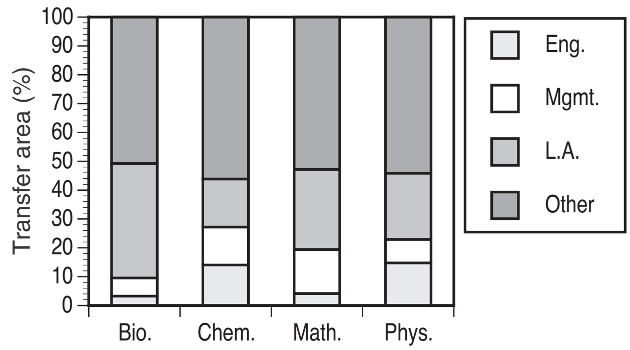
```
            None      Dogs      Cats     Total
   <HS       421        93        28       542
          431.46     73.25     37.29

    HS       666       100        40       806
          641.61    108.93     55.46

   >HS       845       135        99      1079
          858.93    145.82     74.25

 Total      1932       328       167      2427

 ChiSq =   0.253 +   5.326 +   2.316 +
           0.927 +   0.732 +   4.310 +
           0.226 +   0.803 +   8.254 = 23.147
 df = 4, p = 0.000
```

9.44. The graph on the right depicts the conditional distribution of pet ownership for each gender; for example, among females, $\frac{1024}{1266} \doteq 80.88\%$ owned no pets, $\frac{157}{1266} \doteq 12.40\%$ owned dogs, and $\frac{85}{1266} \doteq 6.71\%$ (the rest) owned cats. (One could instead compute column percents—the conditional distribution of gender for each pet-ownership group—but gender makes more sense as the explanatory variable here.) The (slightly altered) Minitab output shows that the relationship between education level and pet ownership is *not* significant ($X^2 \doteq 2.838$, df $= 2$, $P = 0.242$).



Minitab output

```
            None      Dogs      Cats     Total
Female      1024       157        85      1266
         1008.53    170.60     86.86

  Male       915       171        82      1168
          930.47    157.40     80.14

 Total      1939       328       167      2434

 ChiSq =   0.237 +   1.085 +   0.040 +
           0.257 +   1.176 +   0.043 = 2.838
 df = 2, p = 0.242
```



9.45. The missing entries can be seen in the "Other" column of the Minitab output below; they are found by subtracting the engineering, management, and liberal arts counts from each row total. The graph on the right shows the conditional distribution of transfer area for each initial major; for example, of those initially majoring in biology, $\frac{13}{398} \doteq 3.27\%$ transferred to engineering, $\frac{25}{398} \doteq 6.28\%$ transferred to management, and so on. The relationship is significant ($X^2 \doteq 50.53$, df $= 9$, $P < 0.0005$). The largest contributions to X^2 come from chemistry or physics to engineering and biology to liberal arts (more transfers than expected) and biology to engineering and chemistry to liberal arts (fewer transfers than expected).

Minitab output

```
          Eng     Mgmt       LA    Other    Total
  Bio      13       25      158      202      398
        25.30    34.56   130.20   207.95

 Chem      16       15       19       64      114
         7.25     9.90    37.29    59.56

 Math       3       11       20       38       72
         4.58     6.25    23.55    37.62

 Phys       9        5       14       33       61
         3.88     5.30    19.96    31.87

Total      41       56      211      337      645

ChiSq =  5.979 +  2.642 +  5.937 +  0.170 +
        10.574 +  2.630 +  8.973 +  0.331 +
         0.543 +  3.608 +  0.536 +  0.004 +
         6.767 +  0.017 +  1.777 +  0.040 = 50.527
df = 9, p = 0.000
```

9.46. Note that the given counts actually form a three-way table (classified by adhesive, side, and checks). Therefore, this analysis should *not* be done as if the counts come from a 2×4 two-way table; for one thing, no conditional distribution will answer the question of interest (how to avoid face checks). Nonetheless, many students may do this analysis, for which they will find $X^2 = 6.798$, df $= 3$, and $P = 0.079$.

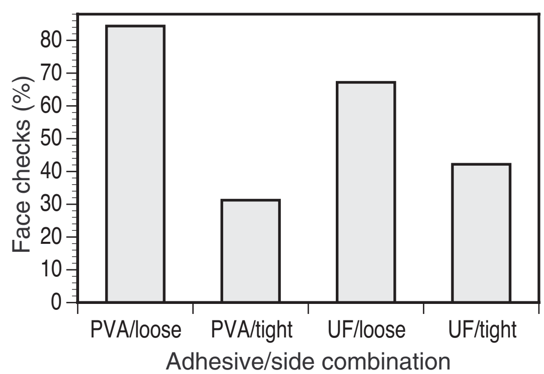
	Face checks	
	No	Yes
PVA/loose	10	54
PVA/tight	44	20
UF/loose	21	43
UF/tight	37	27

A better approach is to rearrange the table as shown on the right. The conditional distributions across the rows will then give us information about avoiding face checks; the graph below illustrates this. We find $X^2 \doteq 45.08$, df $= 3$, and $P < 0.0005$, so we conclude that the appearance of face checks is related to the adhesive/side combination—specifically, we recommend the PVA/tight combination.

Another approach (not quite as good as the previous one) is to perform two separate analyses—say, one for loose side, and one for tight side. These computations show that UF is better than PVA for loose side ($X^2 \doteq 5.151$, df $= 1$, $P = 0.023$), but there is no significant difference for tight side ($X^2 \doteq 1.647$, df $= 1$, $P = 0.200$). We could also do separate analyses for PVA ($X^2 \doteq 37.029$, df $= 1$, $P < 0.0005$) and UF ($X^2 \doteq 8.071$, df $= 1$, $P = 0.005$), from which we conclude that for either adhesive, the tight side has fewer face checks. (Minitab output on the following page.)

Minitab output

```
         NoChk       Chk     Total
PVA-L       10        54        64
         28.00     36.00

PVA-T       44        20        64
         28.00     36.00

 UF-L       21        43        64
         28.00     36.00

 UF-T       37        27        64
         28.00     36.00

Total      112       144       256

ChiSq = 11.571 +  9.000 +
         9.143 +  7.111 +
         1.750 +  1.361 +
         2.893 +  2.250 = 45.079
df = 3, p = 0.000
```



Minitab output

```
-------  Loose side  -------
        NoChk     Chk    Total
  PVA      10      54       64
        15.50   48.50

   UF      21      43       64
        15.50   48.50

Total      31      97      128

ChiSq =  1.952 +  0.624 +
         1.952 +  0.624 = 5.151
df = 1, p = 0.023
-------  Tight side  -------
        NoChk     Chk    Total
  PVA      44      20       64
        40.50   23.50

   UF      37      27       64
        40.50   23.50

Total      81      47      128

ChiSq =  0.302 +  0.521 +
         0.302 +  0.521 = 1.647
df = 1, p = 0.200
```

Minitab output

```
--------  PVA  --------
        NoChk     Chk    Total
Loose      10      54       64
        27.00   37.00

Tight      44      20       64
        27.00   37.00

Total      54      74      128

ChiSq = 10.704 +  7.811 +
        10.704 +  7.811 = 37.029
df = 1, p = 0.000
--------  UF  --------
        NoChk     Chk    Total
Loose      21      43       64
        29.00   35.00

Tight      37      27       64
        29.00   35.00

Total      58      70      128

ChiSq =  2.207 +  1.829 +
         2.207 +  1.829 = 8.071
df = 1, p = 0.005
```

9.47. The Minitab output on the right shows the 2×2 table and significance test details: $X^2 = 852.433$, df $= 1$, $P < 0.0005$. Using $z = -29.2$, computed in the solution to Exercise 8.81(c), this equals z^2 (up to rounding).

Minitab output

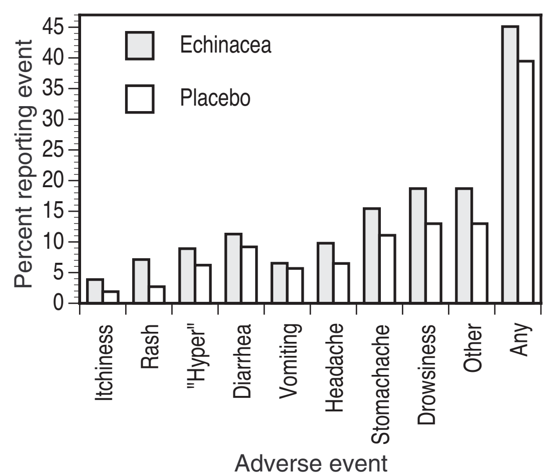
```
          Mex-Am    Other    Total
Juror        339      531      870
          688.25   181.75

  Not     143272    37393   180665
       142922.75 37742.25

Total     143611    37924   181535

ChiSq =177.226 +671.122 +
          0.853 +  3.232 = 852.433
df = 1, p = 0.000
```

9.48. **(a)** The bar graph on the right shows how parental assessment of URIs compares for the two treatments. Note that parental assessment data were apparently not available for all URIs: We have assessments for 329 echinacea URIs and 367 placebo URIs. Minitab output gives $X^2 = 2.506$, df $= 2$, $P = 0.286$, so treatment is not significantly associated with parental assessment. **(b)** If we divide each echinacea count by 337 and each placebo count by 370, we obtain the table of proportions (below, left), and illustrated in the bar graph (below, right). **(c)** The only significant results are for rash ($z = 2.74$, $P = 0.0061$), drowsiness ($z = 2.09$, $P = 0.0366$), and other ($z = 2.09$, $P = 0.0366$). A 10×2 table would not be appropriate, because each URI could have multiple adverse events. **(d)** All results are unfavorable to echinacea, so in this situation we are not concerned that we have falsely concluded that there are differences. In general, when we perform a large number of significance tests and find a few to be significant, we should be concerned that the significant results may simply be due to chance.

Minitab output

```
        Echin  Placebo    Total
Mild      153      170      323
       152.68   170.32

 Mod      128      157      285
       134.72   150.28

 Sev       48       40       88
        41.60    46.40

Total     329      367      696

ChiSq =  0.001 +  0.001 +
         0.335 +  0.300 +
         0.985 +  0.883 = 2.506
df = 2, p = 0.286
```

Event	$\hat{p}_1$	$\hat{p}_2$	z	P
Itchiness	0.0386	0.0189	1.57	0.1154
Rash	0.0712	0.0270	2.74	0.0061
"Hyper"	0.0890	0.0622	1.35	0.1756
Diarrhea	0.1128	0.0919	0.92	0.3595
Vomiting	0.0653	0.0568	0.47	0.6357
Headache	0.0979	0.0649	1.61	0.1068
Stomachache	0.1543	0.1108	1.71	0.0875
Drowsiness	0.1869	0.1297	2.09	0.0367
Other	0.1869	0.1297	2.09	0.0367
Any event	0.4510	0.3946	1.52	0.1290



(e) We would expect multiple observations on the same child to be dependent, so the assumptions for our analysis are not satisfied. Examination of the data reveals that the results for both groups are quite similar, so we are inclined to agree with the authors that there are no statistically significant differences. **(f)** Student opinions about the criticisms of this study will vary. The third criticism might be dismissed as sounding like conspiracy-theory paranoia, but the other three address the way that echinacea was administered; certainly we cannot place too much faith in a clinical trial if it turns out that the treatments were not given properly!

9.49. The chi-square goodness of fit statistic is $X^2 \doteq 3.7807$ with df = 3, for which $P > 0.25$ (software gives 0.2861), so there is not enough evidence to conclude that this university's distribution is different. The details of the computation are given in the table below; note that there were 210 students in the sample.

	Expected frequency	Expected count	Observed count	$O - E$	$\frac{(O-E)^2}{E}$
Never	0.43	90.3	79	−11.3	1.4141
Sometimes	0.35	73.5	83	9.5	1.2279
Often	0.15	31.5	36	4.5	0.6429
Very often	0.07	14.7	12	−2.7	0.4959
			210		3.7807

9.50. The chi-square goodness of fit statistic is $X^2 \doteq 3.4061$ with df = 4, for which $P > 0.25$ (software gives 0.4923), so we have no reason to doubt that the numbers follow a Normal distribution. The details of the computation are given in the table below. The table entries from Table A for −0.6, −0.1, 0.1, and 0.6 are (respectively) 0.2743, 0.4602, 0.5398, and 0.7257. Then, for example, the expected frequency in the interval −0.6 to −0.1 is $0.4602 - 0.2743 = 0.1859$.

	Expected frequency	Expected count	Observed count	$O - E$	$\frac{(O-E)^2}{E}$
$z \le -0.6$	0.2743	137.2	139	1.85	0.0250
$-0.6 < z \le -0.1$	0.1859	93.0	102	9.05	0.8811
$-0.1 < z \le 0.1$	0.0796	39.8	41	1.20	0.0362
$0.1 < z \le 0.6$	0.1859	93.0	78	−14.95	2.4045
$z > 0.6$	0.2743	137.2	140	2.85	0.0592
					3.4061

9.52. The chi-square goodness of fit statistic is $X^2 = 5.50$ with df = 4, for which $0.20 < P < 0.25$ (software gives 0.2397), so we have no reason to doubt that the numbers follow this uniform distribution. The details of the computation are given in the table below.

	Expected frequency	Expected count	Observed count	$O - E$	$\frac{(O-E)^2}{E}$
$0 < x \le 0.2$	0.2	100	114	14	1.96
$0.2 < x \le 0.4$	0.2	100	92	−8	0.64
$0.4 < x \le 0.6$	0.2	100	108	8	0.64
$0.6 < x \le 0.8$	0.2	100	101	1	0.01
$0.8 < x < 1$	0.2	100	85	−15	2.25
					5.50

9.54. A P-value of 0.999 is suspicious because it means that there was an almost-perfect match between the observed and expected counts. (The table on the right shows how small X^2 must be in order to have a P-value of 0.999; recall that X^2 is small when the observed and expected counts are close.) We expect a certain amount of difference between these counts due to chance, and become suspicious if the difference is too small. In particular, when H_0 is true, a match like this would occur only once in 1000 attempts; if there were 1000 students in the class, that might not be too surprising.

df	X^2
1	2×10^{-6}
2	0.0020
3	0.0243
4	0.0908
5	0.2102
6	0.3810
7	0.5985
8	0.8571
9	1.1519
10	1.4787

9.55. (a) Each quadrant accounts for one-fourth of the area, so we expect it to contain one-fourth of the 100 trees. **(b)** *Some* random variation would not surprise us; we no more expect exactly 25 trees per quadrant than we would expect to see exactly 50 heads when flipping a fair coin 100 times. **(c)** The table on the right shows the individual computations, from which we obtain $X^2 = 10.8$, df $= 3$, and $P = 0.0129$. We conclude that the distribution is not random.

Observed	Expected	$(o-e)^2/e$
18	25	1.96
22	25	0.36
39	25	7.84
21	25	0.64
100		10.8

Chapter 10 Solutions

10.1. The given model was $\mu_y = 43.4 + 2.8x$, with standard deviation $\sigma = 4.3$. **(a)** The slope is 2.8. **(b)** When x increases by 1, μ_y increases by 2.8. (Or equivalently, if x increases by 2, μ_y increases by 5.6, etc.) **(c)** When $x = 7$, $\mu_y = 43.4 + 2.8(7) = 63$. **(d)** Approximately 95% of observed responses would fall in the interval $\mu_y \pm 2\sigma = 63 \pm 2(4.3) = 63 \pm 8.6 = 54.4$ to 71.6.

10.2. The regression equation given in Example 10.3 was $\widehat{\text{BMI}} = 29.578 - 0.655\,\text{PA}$. **(a)** With physical activity $x = 9.5$ steps/day, the estimated average BMI is 23.36 kg/m^2. **(b)** The residual is $22.8 - 23.36 \doteq -0.56$. **(c)** Based on the scatterplot (Figure 10.3 in the text), the regression equation is appropriate for 7000 or 12,000 steps/day, but using it for 2000 or 17,000 would be extrapolation, made additionally risky by the suggestion of a curved relationship (Figure 10.5).

10.3. Example 10.5 gives the confidence interval -0.969 to -0.341 for the slope β_1. Recall that slope is the change in y (i.e., $\widehat{\text{BMI}}$) when x (i.e., PA) changes by $+1$. **(a)** If PA increases by 1, we expect $\widehat{\text{BMI}}$ to change by β_1, so the 95% confidence interval for the change is -0.969 to -0.341—that is, a decrease of 0.341 to 0.969 kg/m^2. **(b)** If PA decreases by 1, we expect $\widehat{\text{BMI}}$ to change by $-\beta_1$, so the 95% confidence interval for the change is an increase of 0.341 to 0.969 kg/m^2. **(c)** If PA increases by 0.5, we expect $\widehat{\text{BMI}}$ to change by $0.5\beta_1$, so the 95% confidence interval for the change is a decrease of 0.1705 to 0.4845 kg/m^2.

10.4. The given prediction interval is 16.4 to 31.0 kg/m^2. This interval is $2t^*\text{SE}_{\hat{y}}$ units wide, where $t^* \doteq 1.9845$ for df = 98. Therefore, $\text{SE}_{\hat{y}} \doteq 3.68$. By examining Figure 10.8, we can judge that the prediction intervals for $x = 9$ and $x = 10$ are roughly the same width, so the standard errors should be roughly the same.

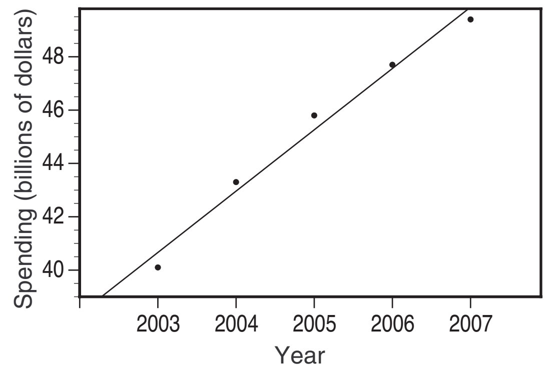
Note: *For the first question, students might take $t^* \doteq 2$, in which case $SE_{\hat{y}} \doteq 3.65$. For the second question, when x changes from 9 to 10, $SE_{\hat{y}}$ increases by about 0.006, so it is quite reasonable to say they are about the same.*

10.5. **(a)** The plot on the following page suggests a linear increase. **(b)** The regression equation is $\hat{y} = -4566.24 + 2.3x$. **(c)** The fitted values and residuals are given in the table on the following page. Squaring the residuals and summing gives 0.952, so the standard error is:

$$s = \sqrt{\frac{0.952}{n-2}} = \sqrt{0.317\overline{3}} \doteq 0.5633$$

(d) Given x (the year), spending comes from a $N(\mu_y, \sigma)$ distribution, where $\mu_y = \beta_0 + \beta_1 x$. The estimates of β_0, β_1, and σ are $b_0 = -4566.24$, $b_1 = 2.3$, and $s \doteq 0.5633$.
(e) We first note that $\bar{x} = 2005$ and $\sum(x_i - \bar{x})^2 = 10$, so $\text{SE}_{b_1} = s/\sqrt{10} \doteq 0.1781$. We have df $= n - 2 = 3$, so $t^* = 3.182$, and the 95% confidence interval for β_1 is $b_1 \pm t^*\text{SE}_{b_1} \doteq 2.3 \pm 0.5667 \doteq 1.733$ to 2.867. This gives the rate of increase of R&D spending: between 1.733 and 2.867 billion dollars per year.

Year	Spending ($billions)	Fitted values	Residuals
2003	40.1	40.66	−0.56
2004	43.3	42.96	0.34
2005	45.8	45.26	0.54
2006	47.7	47.56	0.14
2007	49.4	49.86	−0.46



10.6. **(a)** The variables x and y are reversed: Slope gives the change in y for a change in x. **(b)** The population regression line has intercept β_0 and slope β_1 (not b_0 and b_1). **(c)** The estimate $\hat{\mu}_y = b_0 + b_1x^*$ is more accurate when x^* is close to $\bar{x}$, so the width of the confidence interval grows with $|x^* - \bar{x}|$.

10.7. **(a)** The parameters are β_0, β_1, and σ; b_0, b_1, and s are the *estimates* of those parameters. **(b)** H_0 should refer to β_1 (the population slope) rather than b_1 (the estimated slope). **(c)** The confidence interval will be narrower than the prediction interval because the confidence interval accounts only for the uncertainty in our estimate of the mean response, while the prediction interval must also account for the random error of an individual response.

10.8. The table below gives two sets of answers: those found with critical values from Table D, and those found with software. The approach taken only makes a noticeable difference in part (c), where (with Table D) we take df = 80 rather than df = 98. In each case, the margin of error is $t^*\text{SE}_{b_1} = 0.58t^*$, with df $= n - 2$.

			Table D		Software	
	df	b_1	t^*	Interval	t^*	Interval
(a)	23	1.1	2.069	−0.1000 to 2.3000	2.0687	−0.0998 to 2.2998
(b)	23	2.1	2.069	0.9000 to 3.3000	2.0687	0.9002 to 3.2998
(c)	98	1.1	1.990	−0.0542 to 2.2542	1.9845	−0.0510 to 2.2510

10.9. The test statistic is $t = b_1/\text{SE}_{b_1} = b_1/0.58$, with df $= n - 2$. The tests for parts (a) and (c) are not quite significant at the 5% level, while the test for part (b) is highly significant. This is consistent with the confidence intervals from the previous exercise.

	df	b_1	t	P (Table D)	P (software)
(a)	23	1.1	1.90	$0.05 < P < 0.10$	0.0705
(b)	23	2.1	3.62	$0.001 < P < 0.002$	0.0014
(c)	98	1.1	1.90	$0.05 < P < 0.10$*	0.0608

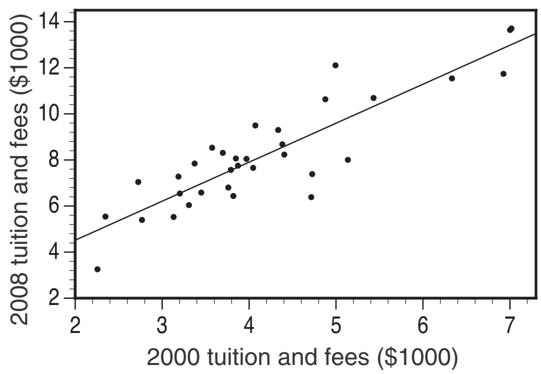
*Note that for (c), if we use Table D, we take df = 80.

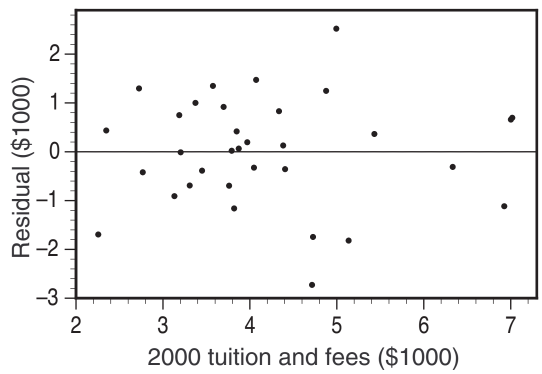
10.10. **(a)** The plot (below, left) shows a strong linear relationship with no striking outliers. **(b)** The regression line (shown on the plot) is $\hat{y} = 1133 + 1.6924x$. **(c)** The residual plot (below, right) shows no clear cause for concern. **(d)** A stemplot (shown) or histogram shows two large residuals (one positive, one negative). **(e)** To test for a relationship, we test H_0: $\beta_1 = 0$ versus H_a: $\beta_1 \neq 0$ (or equivalently, use ρ in place of β_1). **(f)** The test statistic and P-value are given in the Minitab output below: $t \doteq 10.55$, $P < 0.0005$. We have strong evidence of a non-zero slope.

−2	7
−2	
−1	876
−1	11
−0	966
−0	433330
0	0011344
0	66789
1	02234
1	
2	
2	5

Minitab output: Regression of 2008 tuition on 2000 tuition

```
The regression equation is y2008 = 1133 + 1.69 y2000

Predictor        Coef       Stdev    t-ratio        p
Constant       1132.8       701.4       1.61    0.116
y2000          1.6924      0.1604      10.55    0.000

s = 1134        R-sq = 78.2%     R-sq(adj) = 77.5%
```



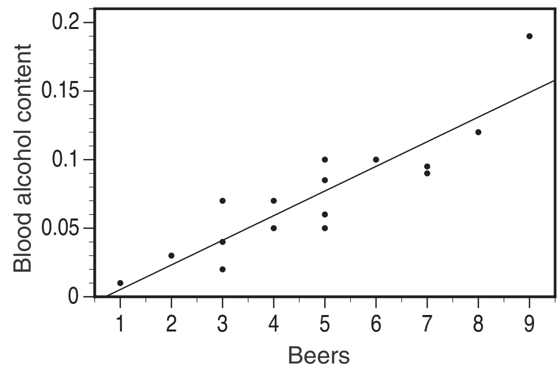


10.11. **(a)** From the Minitab output above, we have $\text{SE}_{b_1} \doteq 0.1604$. With df $= 30$, $t^* = 2.042$, so the 95% confidence interval for β_1 is $1.6924 \pm t^*\text{SE}_{b_1} \doteq 1.3649$ to 2.0199. This slope means that a \$1 difference in tuition in 2000 changes 2008 tuition by between \$1.36 and \$2.02. (It might be easier to understand expressed like this: If the costs of two schools differed by \$1000 in the year 2000, then in 2008, they would differ by between \$1365 and \$2020.) **(b)** Regression explains $r^2 \doteq 78.2\%$ of the variation in 2008 tuition. **(c)** When $x = 5100$, the estimated 2008 tuition is $\hat{y} = 1133 + 1.6924(5100) \doteq \$9764$. **(d)** When $x = 8700$, the estimated 2008 tuition is $\hat{y} = 1133 + 1.6924(8700) \doteq \$15{,}857$. (Software reports \$15,856; the difference is due to rounding. **(e)** The 2000 tuition at Stat U is similar to others in the data set, while Moneypit U was considerably more expensive in 2000, so that prediction requires extrapolation.

Minitab output: Estimates for Stat U and Moneypit U

```
     Fit  Stdev.Fit       95.0% C.I.           95.0% P.I.
    9764        245    (    9264,    10263)  (    7397,    12130)
   15856        749    (   14329,    17384)  (   13084,    18628) X
X denotes a row with very extreme X values
```

10.12. **(a)** β_0 is the population intercept, 0.8. This says that the mean overseas return is 0.8% when the U.S. return is 0%. **(b)** β_1 is the population slope, 0.46. This says that when the U.S. return changes by 1%, the mean overseas return changes by 0.46%. **(c)** The full model is $y_i = 0.8 + 0.46x_i + \epsilon_i$, where y_i and x_i are observed overseas and U.S. returns in a given year, and ϵ_i are independent $N(0, \sigma)$ variables. The residual terms ϵ_i allow for variation in overseas returns when U.S. returns remain the same.



10.13. **(a)** The regression equation is $\hat{y} = -0.0127 + 0.0180x$, and $r^2 \doteq 80.0\%$. Not surprisingly, we find that BAC increases as beer consumption increases; the relationship is quite strong, with beer consumption explaining 80% of the variation in BAC. **(b)** To test H_0: $\beta_1 = 0$ versus H_a: $\beta_1 > 0$, we find $t = 7.48$ and $P < 0.0001$. There is very strong evidence that drinking more beers increases BAC. **(c)** The predicted mean BAC for $x = 5$ beers is 0.07712; the 90% prediction interval is 0.040 to 0.114. Steve might be safe, but cannot be sure that his BAC will be below 0.08.

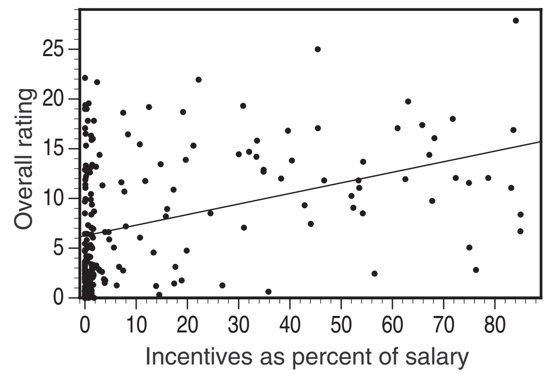
Note: *We use a prediction interval (rather than a confidence interval) because we want a range of values for an* individual *BAC after 5 beers, rather than the* mean *BAC.*

Minitab output: Regression of BAC on beer consumption

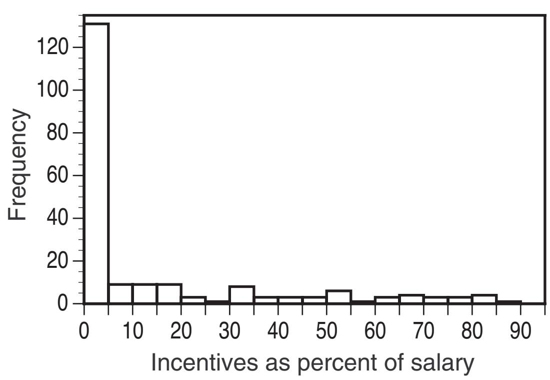
```
The regression equation is BAC = - 0.0127 + 0.0180 Beers

Predictor        Coef       Stdev    t-ratio        p
Constant     -0.01270     0.01264      -1.00    0.332
Beers        0.017964    0.002402       7.48    0.000

s = 0.02044     R-sq = 80.0%     R-sq(adj) = 78.6%

     Fit  Stdev.Fit       90.0% C.I.           90.0% P.I.
 0.07712    0.00513   ( 0.06808, 0.08616)  ( 0.03999, 0.11425)
```

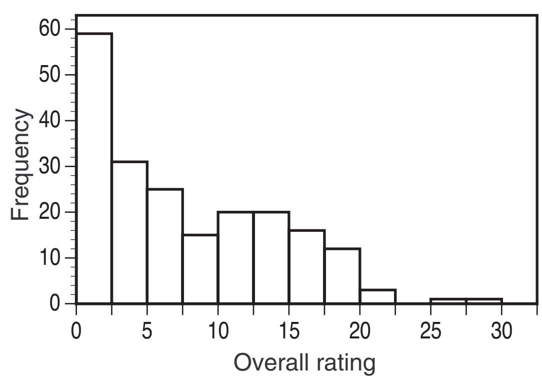
10.14. **(a)** One issue is that correlation coefficients assume a linear association between two variables; if two variables x and y are whole numbers, the only linear relationships between them would have correlation 1 or -1. **(b)** To test H_0: $\rho = 0$ versus H_a: $\rho \neq 0$, we compare $t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$ (with $n = 98$) to a $t(96)$ distribution. The table on the following page shows t and P; entries marked * or ** are significant at the $\alpha = 0.05$ level. **(c)** Entries marked ** are significant at $\frac{0.05}{15} \doteq 0.0033$; that eliminates three significant associations, although two (ASSESS/XTRA and ASSESS/HAND) were barely eliminated, and perhaps should not be completely dismissed. **(d)** We might hesitate to apply the results more broadly because of the artificial setting of these interviews, and the fact that the subjects were undergraduate students.

	XTRA	AGREE	HAND	EYE	DRESS
AGREE	4.53 (0.0000)**				
HAND	2.32 (0.0227)*	0.49 (0.6249)			
EYE	1.79 (0.0761)	1.39 (0.1692)	14.04 (0.0000)**		
DRESS	1.69 (0.0942)	1.08 (0.2809)	4.53 (0.0000)**	4.67 (0.0000)**	
ASSESS	2.86 (0.0052)*	1.28 (0.2020)	2.97 (0.0038)*	3.19 (0.0019)**	1.49 (0.1404)



10.15. (a) Both distributions are sharply right-skewed. Histograms and five-number summaries are below. **(b)** Regression does not require that the variables x and y be Normal; it is the *errors* (the deviation from the line) that should be Normal. **(c)** The scatterplot is on the right; note that incentive pay is the explanatory variable. There is a weak positive linear relationship. **(d)** The regression equation is $\hat{y} = 6.247 + 0.1063x$. (Not surprisingly, regression only explains 15.3% of the variation in rating.) **(e)** Residual analysis might include a histogram or stemplot, a plot of residuals versus incentive pay, and perhaps a Normal quantile plot; the first two of these items are shown on the following page. The residuals are slightly right-skewed. In addition, we note that for incentive pay less than about 30%, most residuals are greater than about -7, but extend up to +15.

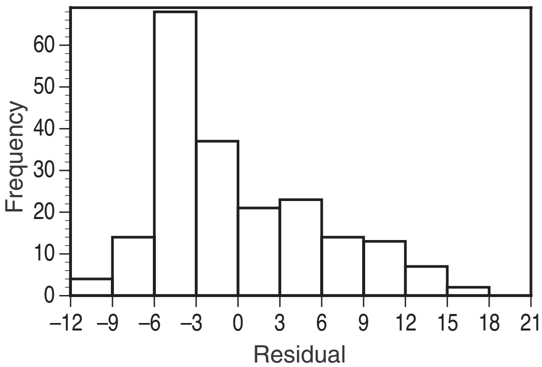


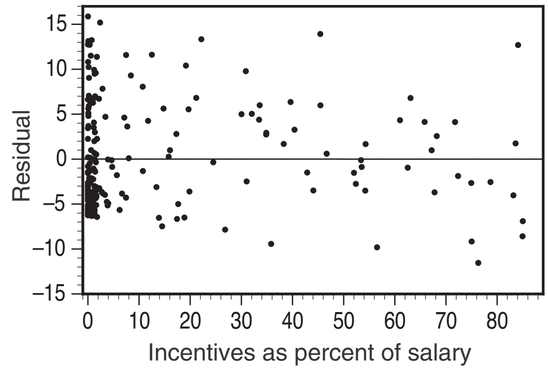


	Min	Q_1	M	Q_3	Max
Incentives as percent of salary	0	0.306	1.43	17.65	85.01
Overall player rating	0	2.25	6.31	12.69	27.88

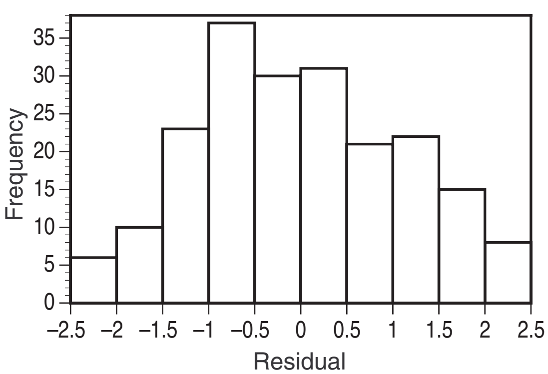
Minitab output: Regression of rating on salary incentive percentage

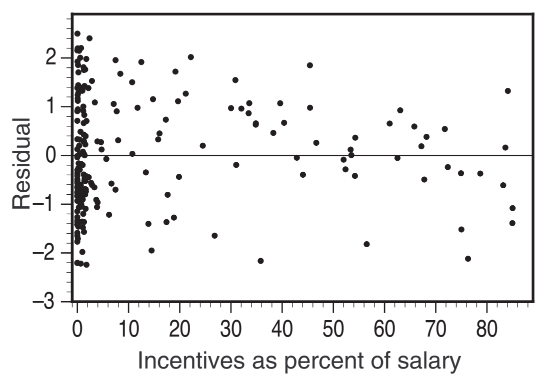
```
The regression equation is Rating = 6.25 + 0.106 Percent

Predictor        Coef       Stdev    t-ratio        p
Constant       6.2469      0.4816      12.97    0.000
Percent       0.10634     0.01767       6.02    0.000

s = 5.854       R-sq = 15.3%     R-sq(adj) = 14.8%
```



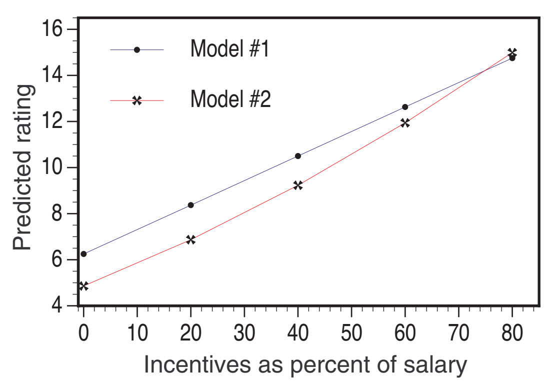


10.16. **(a)** The regression equation is $\hat{y} = 2.2043 + 0.02084x$ (Minitab output on the following page). **(b)** Shown below are a histogram of the residuals, and a plot of residuals versus incentive pay. These show no clear reasons for concern; in particular, the distribution of residuals is considerably less skewed than those in Exercise 10.15. The residual spread is also more consistent for low incentive pay. **(c)** The estimated slope is $b_1 \doteq 0.02084$ with $\text{SE}_{b_1} \doteq 0.003419$. Whether we use df $= 201$ (and software) or df $= 100$ (and Table D), the 95% confidence interval is $b_1 \pm t^* \text{SE}_{b_1} = 0.0141$ to 0.0276. Therefore, if the incentive portion of salary increases by 1%, sqrt(rating) increases between 0.0141 and 0.0246. **(d)** The predicted ratings are shown in the table (below, left). **(e)** In this plot (below, right), we see that the models give similar estimates for high incentive pay, but the square-root model is lower for low incentive pay. **(f)** Based on residuals, this model seems to be better. (There is no clear reason to form a preference based on predicted values.)





Incentive pay	Model #1 (10.15)	Model #2 (10.16)
0%	6.25	4.86
20%	8.37	6.87
40%	10.50	9.23
60%	12.63	11.94
80%	14.75	14.99



Minitab output: Regression of sqrt(rating) on incentive pay

```
The regression equation is SqrtRate = 2.20 + 0.0208 Percent

Predictor       Coef       Stdev    t-ratio        p
Constant     2.20428     0.09321      23.65    0.000
Percent     0.020842    0.003419       6.10    0.000

s = 1.133       R-sq = 15.6%     R-sq(adj) = 15.2%
```

Predicting sqrt(rating) for 0%, 20%, 40%, 60%, 80%

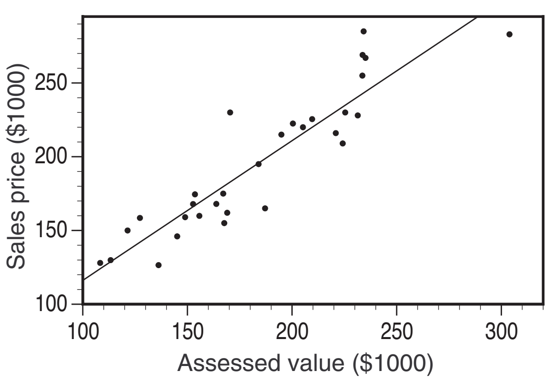
```
    Fit  Stdev.Fit       95.0% C.I.           95.0% P.I.
  2.2043    0.0932   (  2.0205,  2.3881)  ( -0.0377,  4.4463)
  2.6211    0.0819   (  2.4595,  2.7827)  (  0.3808,  4.8614)
  3.0380    0.1187   (  2.8038,  3.2721)  (  0.7913,  5.2847)
  3.4548    0.1756   (  3.1085,  3.8011)  (  1.1937,  5.7159)
  3.8716    0.2386   (  3.4011,  4.3422)  (  1.5882,  6.1551)
```

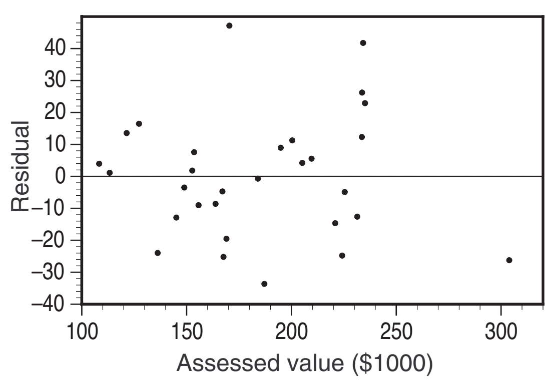
Predicting rating for 0%, 20%, 40%, 60%, 80%

```
    Fit  Stdev.Fit       95.0% C.I.           95.0% P.I.
  6.247     0.482    (   5.297,   7.197)  (  -5.337,  17.831)
  8.374     0.423    (   7.539,   9.209)  (  -3.201,  19.949)
 10.500     0.613    (   9.291,  11.710)  (  -1.108,  22.108)
 12.627     0.907    (  10.838,  14.416)  (   0.945,  24.310)
 14.754     1.233    (  12.323,  17.185)  (   2.956,  26.552)
```

10.17. **(a)** 22 of these 30 homes sold for more than their assessed values. This was "an SRS of 30 properties," so it should be reasonably representative, so the larger population should be similar (or at least it was at the time of the sample). **(b)** The scatterplot (below, left) shows a moderately strong linear association. **(c)** The regression line $\hat{y} = 21.50 + 0.9468x$ is included on the scatterplot. **(d)** The plot of residuals versus assessed value (below, right) shows no obvious unusual features. The house with the highest assessed value (which also stands out in the original scatterplot) may be influential. **(e)** A stemplot (shown here) or histogram looks reasonably Normal, although there are two high residuals that stand apart from the rest. **(f)** There are no clear violations of the assumptions—at least, none severe enough to cause too much concern.

−3	3
−2	6543
−1	9422
−0	984430
0	1134578
1	1236
2	26
3	
4	17





Minitab output: Regression of sales price on assessed value

```
The regression equation is SalesPrc = 21.5 + 0.947 Assessed

Predictor       Coef       Stdev    t-ratio        p
Constant       21.50       15.28       1.41    0.170
Assessed     0.94682     0.08064      11.74    0.000

s = 19.73       R-sq = 83.1%     R-sq(adj) = 82.5%
```

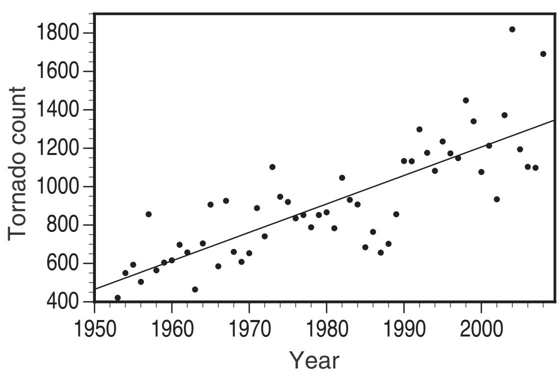
10.18. (a) and **(b)** The table on the right gives the predicted selling prices and residuals for these three houses. **(c)** From the Minitab output in the previous solution, we have $b_0 \doteq 21.50$, $\text{SE}_{b_0} \doteq 15.28$, $b_1 \doteq 0.94682$, and $\text{SE}_{b_1} \doteq 0.08064$. For 95% confidence with df $= 28$, we have $t^* = 2.048$, so the intervals are -9.79 to 52.79 (intercept) and 0.78 to 1.11 (slope). **(d)** The two confidence intervals in part (c) include the values 0 (for the intercept) and 1 (for the slope), so we cannot reject $y = x$ as unreasonable.

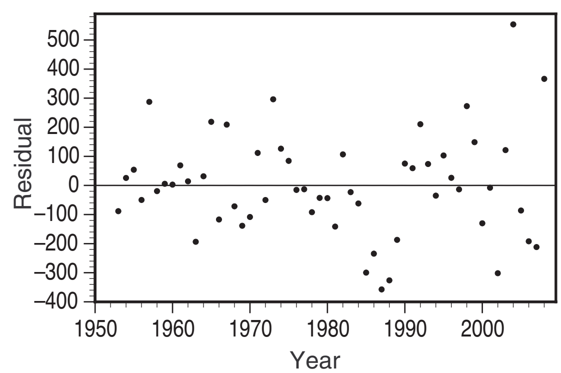
Assessed value	Selling price	Predicted price	Residual
155	142.9	168.26	−25.36
220	224.0	229.80	−5.80
285	286.0	291.34	−5.34

Note: *In Exercise 10.17(a), we noted that 22 of the 30 homes in the sample sold for more than their assessed value, and some students might fall back on that to answer part (d) of this question. However, the confidence intervals in part (c) suggest that we do not have enough evidence to reject the null hypothesis that the model is* $y = x$.

10.19. (a) The plot (below, left) is roughly linear and increasing. The number of tornadoes in 2004 (1819) is noticeably high, as is the 2008 count (1691) to a lesser extent. **(b)** The regression equation is $\hat{y} \doteq -28{,}438 + 14.82x$; both the slope and intercept are significantly different from 0. In the Minitab output on the following page, we see $\text{SE}_{b_1} \doteq 1.463$. With $t^* = 2.0049$ for df $= 54$, the confidence interval for β_1 is $b_1 \pm t^*\text{SE}_{b_1} = 14.82 \pm 2.93 \doteq 11.89$ to 17.76 tornadoes per year. **(c)** In the plot (below, right), we see that the scatter might be greater in recent years, and the 2004 residual is particularly high. **(d)** Based on a stemplot (right), the 2004 residual is an outlier; the other residuals appear to be roughly Normal. **(e)** Without the 2004 count, the regression equation is $\hat{y} \doteq -26{,}584 + 13.88x$. The estimated slope decreases by almost one tornado per year.

−3	520
−2	931
−1	99843310
−0	9887654443211110
0	001223556778
1	001224
2	011789
3	6
4	
5	5





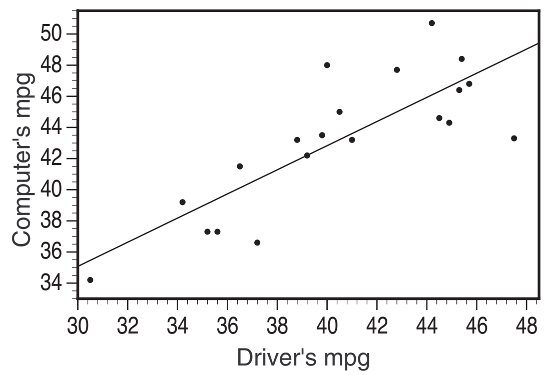
Minitab output: Regression of tornado count on year

```
The regression equation is Count = - 28438 + 14.8 Year

Predictor       Coef       Stdev    t-ratio        p
Constant      -28438        2897      -9.82    0.000
Year          14.822       1.463      10.13    0.000

s = 176.9       R-sq = 65.5%     R-sq(adj) = 64.9%
```

Regression with 2004 count removed

```
The regression equation is Count2 = - 26584 + 13.9 Year2

Predictor       Coef       Stdev    t-ratio        p
Constant      -26584        2680      -9.92    0.000
Year2         13.881       1.354      10.26    0.000

s = 160.5       R-sq = 66.5%     R-sq(adj) = 65.9%
```

10.20. **(a)** The scatterplot shows a fairly strong positive linear association, with no extreme outliers, so regression seems to be appropriate. **(b)** The regression equation (shown on the scatterplot) is $\hat{y} = 11.81 + 0.7754x$. **(c)** Student summaries will vary. The mpg values are certainly similar, but one notable difference is that all but three of the computer values are higher than the driver's values, and the mean computer mpg is about 2.7 mpg higher than the mean driver mpg. Additionally, the slope of the regression line is about 0.78, meaning that (on average) a 1 mpg change in the driver's value corresponds to a 0.78 mpg for the computer. The intercept, however, is about 11.8 mpg, suggesting that the computer's value is generally higher when the driver's value is small.



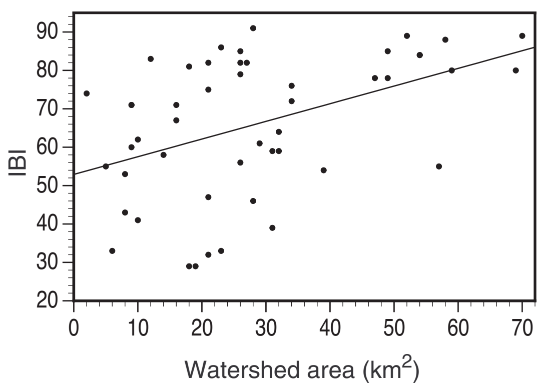
Minitab output: Regression of computer mpg on driver mpg

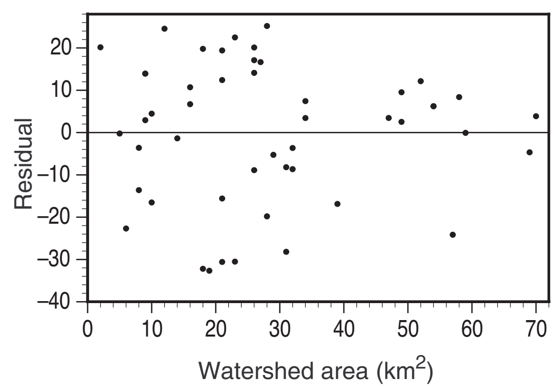
```
The regression equation is Computer = 11.8 + 0.775 Driver

Predictor       Coef       Stdev    t-ratio        p
Constant      11.812       5.432       2.17    0.043
Driver        0.7754      0.1335       5.81    0.000

s = 2.676       R-sq = 65.2%     R-sq(adj) = 63.3%
```

10.21. **(a)** About $r^2 \doteq 8.41\%$ of the variability in AUDIT score is explained by (a linear regression on) gambling frequency. **(b)** With $r = 0.29$ and $n = 908$, the test statistic for H_0: $\rho = 0$ versus H_a: $\rho \neq 0$ is $t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \doteq 9.12$ (df $= 906$), for which P is very small. **(c)** Nonresponse is a problem because the students who did not answer might have different characteristics from those who did. Because of this, we should be cautious about considering these results to be representative of all first-year students at this university, and even more cautious about extending these results to the broader population of all first-year students.

10.22. **(a)** Stemplots are shown on the right. x (watershed area) is right-skewed; $\bar{x} \doteq 28.2857$ km^2, $s_x \doteq 17.7142$ km^2. y (IBI) is left-skewed; $\bar{y} \doteq 65.9388$, $s_y \doteq 18.2796$. **(b)** The scatterplot (below, left) shows a weak positive association, with more scatter in y for small x. **(c)** $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, $i = 1, 2, ..., 49$; ϵ_i are independent $N(0, \sigma)$ variables. **(d)** The hypotheses are H_0: $\beta_1 = 0$ versus H_a: $\beta_1 \neq 0$. **(e)** See the Minitab output below. The regression equation is $\widehat{\text{IBI}} = 52.92 + 0.4602$ Area, and the estimated standard deviation is $s \doteq 16.53$. For testing the hypotheses in part (d), $t = 3.42$ and $P = 0.001$. **(f)** The residual plot (below, right) again shows that there is more variation for small x. **(g)** As we can see from a stemplot and/or a Normal quantile plot (both below), the residuals are somewhat left-skewed but otherwise seem reasonably close to Normal. **(h)** Student opinions may vary. The two apparent deviations from the model are (i) a possible change in standard deviation as x changes and (ii) possible non-Normality of error terms.

Area		IBI	
0	2	2	99
0	5688999	3	233
1	0024	3	9
1	66889	4	13
2	111133	4	67
2	66667889	5	34
3	112244	5	556899
3	9	6	0124
4		6	7
4	799	7	11124
5	244	7	56889
5	789	8	001222344
6		8	556899
6	9	9	1
7	0		

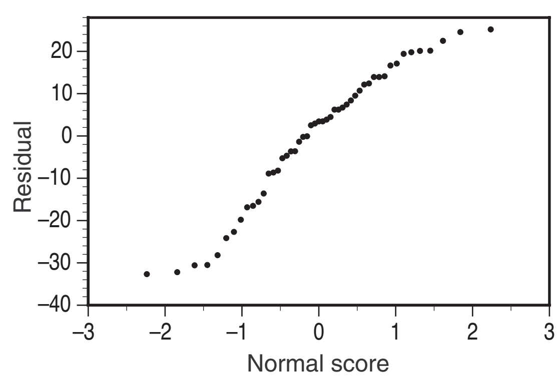




Minitab output: Regression of IBI on watershed area

```
The regression equation is IBI = 52.9 + 0.460 Area

Predictor        Coef        Stdev    t-ratio        p
Constant       52.923        4.484      11.80    0.000
Area           0.4602       0.1347       3.42    0.001

s = 16.53       R-sq = 19.9%     R-sq(adj) = 18.2%
```

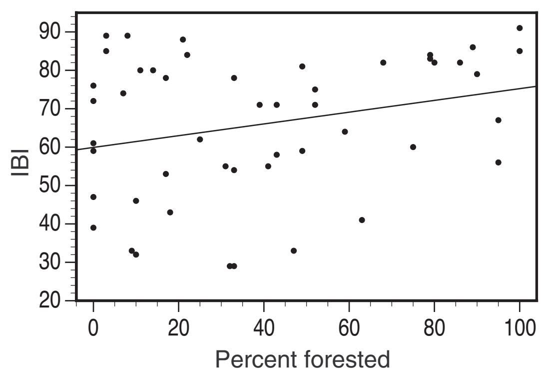
−3	2200
−2	8
−2	42
−1	9665
−1	3
−0	8885
−0	433100
0	223334
0	666789
1	022334
1	6799
2	0024
2	5

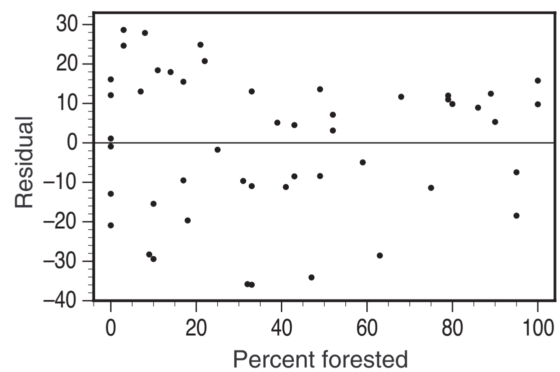


10.23. (a) The stemplot of percent forested is shown on the right; see the solution to the previous exercise for the stemplot of IBI. x (percent forested) is right-skewed; $\bar{x} = 39.3878\%$, $s_x = 32.2043\%$. y (IBI) is left-skewed; $\bar{y} = 65.9388$, $s_y = 18.2796$. **(b)** The scatterplot (below, left) shows a weak positive association, with more scatter in y for small x. **(c)** $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, $i = 1, 2, ..., 49$; ϵ_i are independent $N(0, \sigma)$ variables. **(d)** The hypotheses are H_0: $\beta_1 = 0$ versus H_a: $\beta_1 \neq 0$. **(e)** See the Minitab output below. The regression equation is $\widehat{\text{IBI}} = 59.91 + 0.1531$ Forest, and the estimated standard deviation is $s \doteq 17.79$. For testing the hypotheses in (d), $t = 1.92$ and $P = 0.061$. **(f)** The residual plot (below, right) shows a slight curve—the residuals seem to be (very) slightly lower in the middle and higher on the ends. **(g)** As we can see from a stemplot and/or a Normal quantile plot (both below), the residuals are left-skewed. **(h)** Student opinions may vary. The three apparent deviations from the model are (i) a possible change in standard deviation as x changes, (ii) possible curvature of residuals, and (iii) possible non-Normality of error terms.

Percent forested

0	00000033789
1	0014778
2	125
3	123339
4	133799
5	229
6	38
7	599
8	069
9	055
10	00

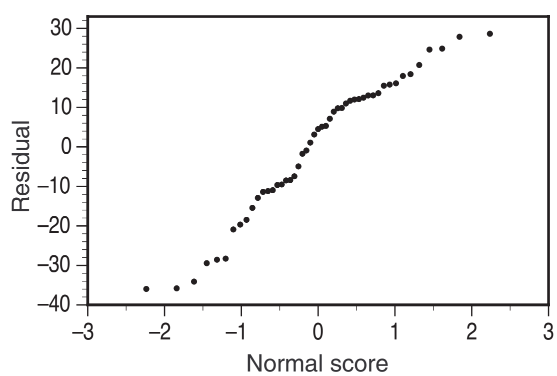




Minitab output: Regression of IBI on percent forested

```
The regression equation is IBI = 59.9 + 0.153 Forest

Predictor        Coef        Stdev     t-ratio        p
Constant       59.907        4.040       14.83    0.000
Forest        0.15313      0.07972        1.92    0.061

s = 17.79       R-sq = 7.3%      R-sq(adj) = 5.3%
```

−3	55
−3	4
−2	988
−2	0
−1	985
−1	2110
−0	99887
−0	410
0	134
0	557899
1	01122333
1	55678
2	044
2	78



10.24. The first model (using watershed area to predict IBI) is preferable because the regression was significant ($P = 0.001$ versus $P = 0.061$) and explained a higher proportion of the variation in IBI (19.9% versus 7.3%).

10.25. The precise results of these changes depend on which observation is changed. (There are six observations which had 0% forest and two which had 100% forest.) Specifically, if we change IBI to 0 for one of the first six observations, the resulting P-value is between 0.019 (observation 6) and 0.041 (observation 3). Changing one of the last two observations changes the P-value to 0.592 (observation 48) or 0.645 (observation 49).

In general, the first change decreases P (that is, the relationship is more significant) because it accentuates the positive association. The second change weakens the association, so P increases (the relationship is less significant).

10.26. With the regression equation $\widehat{\text{IBI}} = 52.92 + 0.4602\ \text{Area}$, the predicted mean response when $x = \text{Area} = 40\ \text{km}^2$ is $\hat{\mu}_y = \widehat{\text{IBI}} \doteq 71.33$. While it is possible to find $\text{SE}_{\hat{\mu}}$ and $\text{SE}_{\hat{y}}$ using the formulas from Section 10.2, we rely on the software output shown below. ($\text{SE}_{\hat{\mu}} \doteq 2.84$, reported by Minitab as "Stdev.fit," and $\text{SE}_{\hat{y}} = \sqrt{s^2 + \text{SE}_{\hat{\mu}}^2} \doteq 16.77$, where $s \doteq 16.53$ was given in the Minitab output shown with the solution to Exercise 10.22. For df $= 47$, the appropriate critical value is $t^* = 2.0117$.) **(a)** The 95% confidence interval for μ_y is 65.61 to 77.05. **(b)** The 95% prediction interval for a future response is 37.57 to 105.09. **(c)** Among *many* streams with watershed area 40 km^2, we estimate the mean IBI to be between about 65.61 and 77.05. For an *individual* stream with watershed area 40 km^2, we expect its IBI to be between about 37.57 and 105.09. **(d)** We probably cannot reliably apply these results elsewhere; it is likely that the particular characteristics of the Ozark Highland region play some role in determining the regression coefficients.

Minitab output: Predicting IBI for watershed area = 40 km^2

```
   Fit  Stdev.Fit       95.0% C.I.           95.0% P.I.
 71.33       2.84   (   65.61,   77.05)  (   37.57,  105.09)
```

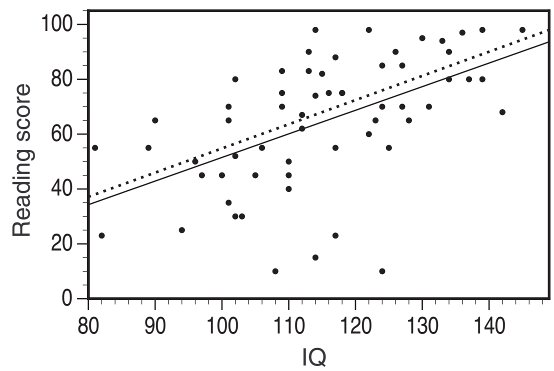
10.27. Using $\text{Area} = 10$ in the model $\widehat{\text{IBI}} = 52.92 + 0.4602\ \text{Area}$ from Exercise 10.22, $\widehat{\text{IBI}} \doteq 57.52$. Using $\text{Forest} = 63$ in the model $\widehat{\text{IBI}} = 59.91 + 0.1531\ \text{Forest}$ from Exercise 10.23, $\widehat{\text{IBI}} \doteq 69.55$. Both predictions have a lot of uncertainty; recall that r^2 was fairly small for both models. Also note that the prediction intervals (shown below) are both about 70 units wide.

Minitab output: Predicting IBI for watershed area = 10

```
   Fit  Stdev.Fit       95.0% C.I.           95.0% P.I.
 57.52       3.41   (   50.66,   64.39)  (   23.55,   91.50)
```

Predicting IBI for percent forest = 63

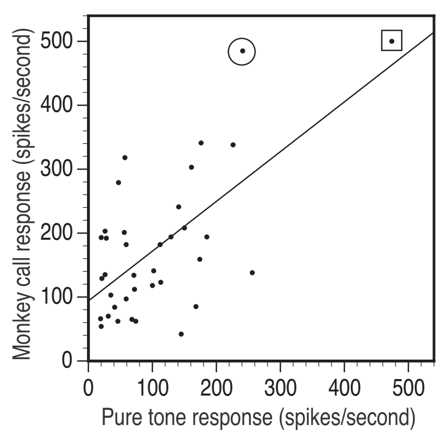
```
   Fit  Stdev.Fit       95.0% C.I.           95.0% P.I.
 69.55       3.16   (   63.19,   75.92)  (   33.20,  105.91)
```



10.28. **(a)** With all 60 points, the regression equation is $\hat{y} = -34.55 + 0.8605x$, $s \doteq 20.17$. (This is the solid line in the scatterplot on the right.) The slope is significantly different from 0: $t = 4.82$, $P < 0.0005$. **(b)** Without the four points from the bottom of the scatterplot, the regression equation is $\hat{y} = -33.40 + 0.8818x$, $s \doteq 15.18$. (This is the dashed line in the scatterplot.) The slope is again significantly different from 0: $t = 6.57$, $P < 0.0005$. With the outliers removed, the line changes slightly; the most significant change is the decrease in the estimated standard deviation s. This correspondingly makes t larger (i.e., b_1 is *more* significantly different from 0) and makes the regression line more useful for prediction (r^2 increases from 28.9% to 44.4%). Of course, we should not arbitrarily remove data points; more investigation is needed to determine why these students' reading scores were so much lower than we would expect based on their IQs.

10.29. **(a)** Stemplots are shown below; both variables are right-skewed. For pure tones, $\bar{x} \doteq 106.20$ and $s \doteq 91.76$ spikes/second, and for monkey calls, $\bar{y} = 176.57$ and $s_y = 111.85$ spikes/second. **(b)** There is a moderate positive association; the third point (circled) has the largest residual; the first point (marked with a square) is an outlier for tone response. **(c)** With all 37 points, $\widehat{\text{CALL}} = 93.9 + 0.778\ \text{TONE}$ and $s = 87.30$; the test of $\beta_1 = 0$ gives $t = 4.91$, $P < 0.0001$. **(d)** Without the first point, $\hat{y} = 101 + 0.693x$, $s = 88.14$, $t = 3.18$. Without the third point, $\hat{y} = 98.4 + 0.679x$, $s = 80.69$, $t = 4.49$. With neither, $\hat{y} = 116 + 0.466x$, $s = 79.46$, $t = 2.21$. The line changes a bit, but always has a slope significantly different from 0.

Tone		Call		Residual	
0	122222233444	0	4	−1	65
0	55556777	0	566667889	−1	3
1	0011244	1	011223334	−0	8876555
1	566778	1	5889999	−0	44444331100
2	24	2	0004	0	012334
2	5	2	7	0	667888
3		3	0134	1	14
3		3		1	7
4		4		2	0
4	7	4	8		
		5	0		



10.30. The model is $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$; ϵ_i are independent $N(0, \sigma)$ random variables. **(a)** β_0 represents the fixed costs. **(b)** β_1 represents how costs change as the number of students changes. This should be positive because more students mean more expenses. **(c)** The error term (ϵ_i) allows for variation among equal-sized schools.

10.31. (a) The stemplots (below, left) are fairly symmetric. For x (MOE), $\bar{x} \doteq 1,799,180$ and $s_x \doteq 329,253$; for y (MOR), $\bar{y} \doteq 11,185$ and $s_y \doteq 1980$. **(b)** The plot (below, right) shows a moderately strong, positive, linear relationship. Because we would like to predict MOR from MOE, we should put MOE on the x axis. **(c)** The model is $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, $i = 1, 2, ..., 32$; ϵ_i are independent $N(0, \sigma)$ variables. The regression equation is $\widehat{\text{MOR}} = 2653 + 0.004742\,\text{MOE}$, $s \doteq 1238$. The slope is significantly different from 0: $t = 7.02$ (df = 30), $P < 0.0001$. **(d)** Assumptions appear to be met: A stemplot of the residuals shows one slightly low (not quite an outlier), but acceptable, and the plot of residuals against MOE (not shown) does not suggest any particular pattern.

MOE

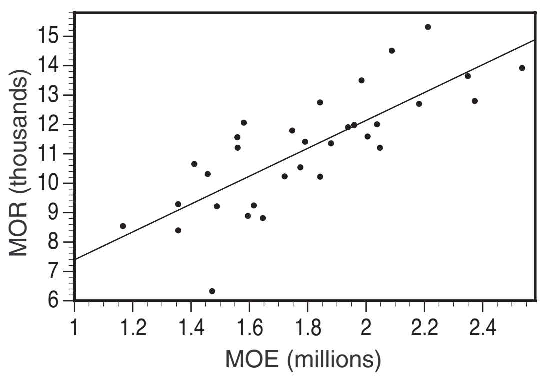
11	6
12	
13	55
14	1578
15	5589
16	14
17	2479
18	447
19	358
20	0348
21	8
22	1
23	47
24	
25	3

MOR

6	3
7	
8	3588
9	222
10	22356
11	223455799
12	00777
13	469
14	5
15	3

Residuals

−3	3
−2	
−2	
−1	6
−1	31110
−0	76555
−0	43221
0	00223
0	78
1	1334
1	599
2	1

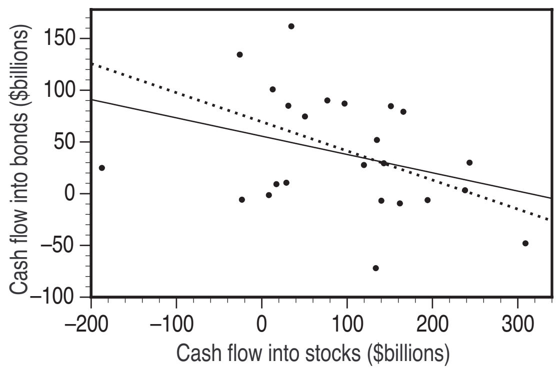


10.32. (a) The 95% confidence interval gives a range of values for the mean MOR of *many* pieces of wood with MOE equal to 2,400,000. The prediction interval gives a range of values for the MOR of *one* piece of wood with MOE equal to 2,400,000. **(b)** The prediction interval will include more values because the confidence interval accounts only for the uncertainty in our estimate of the mean response, while the prediction interval must also account for the random error of an individual response. **(c)** With the regression equation $\widehat{\text{MOR}} = 2653 + 0.004742\,\text{MOE}$, the predicted mean response when $x = \text{MOE} = 2{,}400{,}000$ is $\hat{\mu}_y = \widehat{\text{MOE}} \doteq 14{,}034$. The Minitab output below gives the two intervals, along with $\text{SE}_{\hat{\mu}}$ ("Stdev.fit").

Minitab output: Predicting MOR with MOE = 2,400,000

```
   Fit  Stdev.Fit       95.0% C.I.          95.0% P.I.
 14034        461   (   13092,   14976)  (   11335,   16733)
```

10.33. **(a)** The scatterplot shows a weak negative association; the regression equation is $\widehat{\text{Bonds}} = 55.58 - 0.1769\ \text{Stocks}$, with $s \doteq 54.55$. (This is the solid line in the plot.) **(b)** For testing H_0: $\beta_1 = 0$ versus H_a: $\beta_1 \neq 0$, we have $t = -1.66$ (df $= 22$) and $P = 0.111$. The slope is *not* significantly different from 0. **(c)** With the 2008 data removed, the (dashed) regression line is $\widehat{\text{Bonds}} = 69.46 - 0.2814\ \text{Stocks}$, with $s \doteq 53.12$. The slope is now significantly different from 0 ($t = -2.24$, $P = 0.036$). **(d)** We should explore whether something happened in 2008 that might explain why that point strayed from the line. (The economy would seem to be a likely cause.)



Minitab output: Regression of bond cash flow on stock cash flow

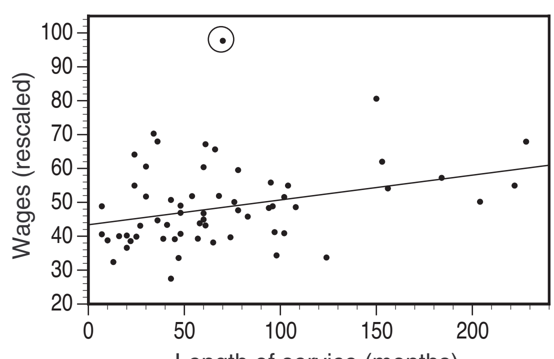
```
The regression equation is Bonds = 55.6 - 0.177 Stocks

Predictor      Coef      Stdev    t-ratio       p
Constant      55.58      14.98       3.71   0.001
Stocks      -0.1769     0.1066      -1.66   0.111

s = 54.55      R-sq = 11.1%    R-sq(adj) = 7.1%
```

Regression without 2008 data

```
The regression equation is Bonds = 69.5 - 0.281 Stocks

Predictor      Coef      Stdev    t-ratio       p
Constant      69.46      17.33       4.01   0.001
Stocks      -0.2814     0.1254      -2.24   0.036

s = 53.12      R-sq = 19.3%    R-sq(adj) = 15.5%
```

10.34. **(a)** The t statistic for testing H_0: $\beta_1 = 0$ versus H_a: $\beta_1 \neq 0$ is $t = b_1/\text{SE}_{b_1} = 0.72/0.38 \doteq 1.89$ with df $= 80$. This has $P = 0.0617$, so we do not reject H_0. **(b)** For the one-sided alternative $\beta_1 > 0$, we would have $P = 0.0309$, so we could reject H_0 at the 5% significance level.

10.35. **(a)** Aside from the one high point (70 months of service, and wages 97.6801), there is a moderate positive association—fairly clear but with quite a bit of scatter. **(b)** The regression equation is $\widehat{\text{WAGES}} = 43.383 + 0.07325\ \text{LOS}$, with $s \doteq 10.21$ (Minitab output follows). The slope is significantly different from 0: $t = 2.85$ (df $= 57$), $P = 0.006$. **(c)** Wages rise an average of 0.07325 wage units per week of service. **(d)** We have $b_1 \doteq 0.07325$ and $\text{SE}_{b_1} \doteq 0.02571$. For a t distribution with df $= 57$, $t^* \doteq 2.0025$ for a 95% confidence interval, so the interval is 0.0218 to 0.1247.



Minitab output: Regression of wages on length of service (outlier excluded)

```
The regression equation is wages = 43.4 + 0.0733 los

Predictor        Coef        Stdev    t-ratio        p
Constant       43.383        2.248      19.30    0.000
los           0.07325      0.02571       2.85    0.006

s = 10.21       R-sq = 12.5%     R-sq(adj) = 10.9%
```

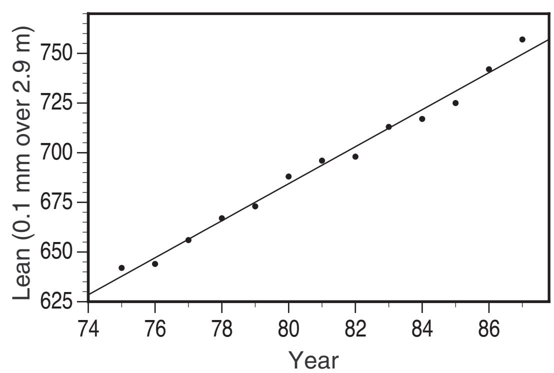
Regression of wages on length of service (outlier included)

```
The regression equation is wages = 44.2 + 0.0731 los

Predictor        Coef        Stdev    t-ratio        p
Constant       44.213        2.628      16.82    0.000
los           0.07310      0.03015       2.42    0.018

s = 11.98       R-sq = 9.2%      R-sq(adj) = 7.6%
```

10.36. The table below summarizes the regression results with the outlier excluded, and those with all points. Minitab output for both regressions is shown above. **(a)** The intercept and slope estimates change very little, but the estimate of σ increases from 10.21 to 11.98. **(b)** With the outlier, the t statistic decreases (because s has increased), and the P-value increases slightly—although it is still significant at the 5% level. **(c)** The interval width $2t^*\mathrm{SE}_{b_1}$ increases from 0.1030 to 0.1207—roughly the same factor by which s increased. (Because the degrees of freedom change from 57 to 58, t^* decreases from 2.0025 to 2.0017, but the change in s has a much greater impact.)

	b_0	b_1	s	t	P	Interval width
Outlier excluded	43.383	0.07325	10.21	2.85	0.006	0.1030
All points	44.213	0.07310	11.98	2.42	0.018	0.1207

10.37. **(a)** The trend appears to be quite linear. **(b)** The regression equation is $\widehat{\text{Lean}} = -61.12 + 9.3187\,\text{Year}$ with $s \doteq 4.181$. The regression explains $r^2 = 98.8\%$ of the variation in lean. **(c)** The rate we seek is the slope. For df $= 11$ and 99% confidence, $t^* = 3.1058$, so the interval is $9.3187 \pm (3.1058)(0.3099) = 8.3562$ to 10.2812 tenths of a millimeter/year.



Minitab output: Regression of lean on year

```
The regression equation is Lean = -61.1 + 9.32 Year

Predictor        Coef        Stdev    t-ratio        p
Constant       -61.12        25.13      -2.43    0.033
Year           9.3187       0.3099      30.07    0.000

s = 4.181       R-sq = 98.8%     R-sq(adj) = 98.7%
```

10.38. **(a)** $\hat{y} = -61.12 + 9.3187(18) \doteq 107$, for a prediction of 2.9107 m. **(b)** This is an example of extrapolation—trying to make a prediction outside the range of given x-values. Minitab reports that a 95% prediction interval for $\hat{y}$ when $x^* = 18$ is about 62.6 to 150.7. The width of the interval is an indication of how unreliable the prediction is.

Note: *Minitab's "Stdev.Fit" value of 19.56 is $SE_{\hat{\mu}}$, so $SE_{\hat{y}} \doteq \sqrt{s^2 + SE_{\hat{\mu}}^2} \doteq 20.00$, which agrees with the margin for the prediction interval: $t^* SE_{\hat{y}} = (2.201)(20.00) \doteq 44.02$.*

Minitab output: Predicting lean in 1918 (year = 18)

```
   Fit  Stdev.Fit       95.0% C.I.          95.0% P.I.
106.62      19.56   (   63.56,  149.68)  (   62.58,  150.65) XX
XX denotes a row with very extreme X values
```

10.39. **(a)** Use $x = 112$ (the number of years after 1900). **(b)** $\hat{y} = -61.12 + 9.3187(112) \doteq 983$, for a prediction of 2.9983 m. **(c)** A prediction interval is appropriate because we are interested in one future observation, not the mean of all future observations; in this situation, it does not make sense to talk of more than one future observation. In the output below, note that Minitab warns us of the risk of extrapolation.

Minitab output: Predicting lean in 2012 (year = 112)

```
   Fit  Stdev.Fit       95.0% C.I.          95.0% P.I.
982.57       9.68   (  961.27, 1003.88)  (  959.36, 1005.78) XX
XX denotes a row with very extreme X values
```

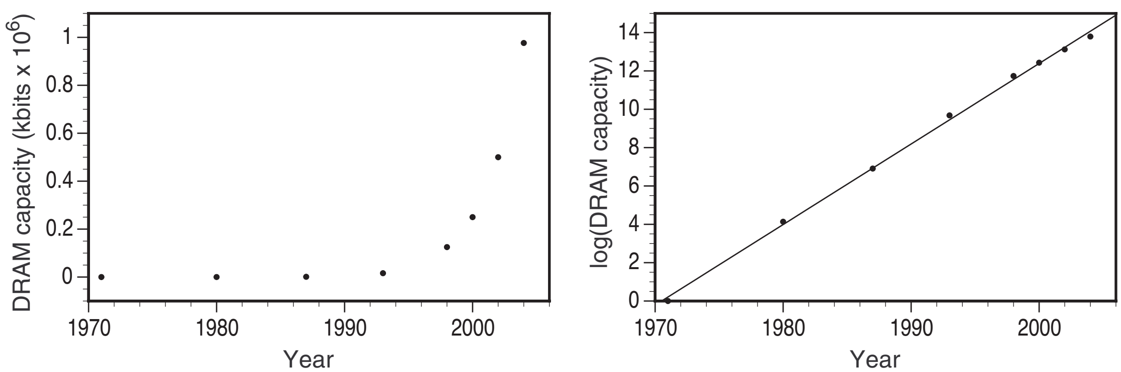
10.40. A negative association makes sense here: If the price of beer is above average, fewer students can afford to drink, while more drinking happens when beer is cheaper.

Note: *The fact that the correlation is relatively small indicates that the price of beer is not a crucial factor in determining the prevalence of binge-drinking. In particular, a straight-line relationship with the cost of beer only explains about $r^2 \doteq 13\%$ of the variation in binge-drinking rates.*

10.41. To test H_0: $\rho = 0$ versus H_a: $\rho \neq 0$, we compute $t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \doteq -4.16$. Comparing this to a t distribution with df $= 116$, we find $P < 0.0001$, so we conclude the correlation is different from 0.

10.42. **(a)** Scatterplot below, left. **(b)** Scatterplot below, right. **(c)** The regression equation is $\hat{y} = -827.66 + 0.4200x$ with $s \doteq 0.2016$. For 95% confidence with df $= 6$, $t^* \doteq 2.4469$, so with $b_1 \doteq 0.4200$ and $\text{SE}_{b_1} \doteq 0.006524$, the confidence interval is 0.4041 to 0.4360.

Note: *If students use a common logarithm (rather than a natural logarithm, as we have done), everything would be multiplied by about 0.4343: The vertical scale on the graph would be from 0 to about 6, the regression line would be* $\hat{y} = -359.45 + 0.1824x$, *and the interval would be 0.1755 to 0.1893.*



Minitab output: Regression of log(kilobits) on year

```
The regression equation is logKbits = - 828 + 0.420 Year

Predictor        Coef       Stdev    t-ratio        p
Constant      -827.66       12.99     -63.69    0.000
Year         0.420024    0.006524      64.38    0.000

s = 0.2016      R-sq = 99.9%     R-sq(adj) = 99.8%
```

10.43. Recall that testing H_0: $\rho = 0$ versus H_a: $\rho \neq 0$ is the same as testing H_0: $\beta_1 = 0$ versus H_a: $\beta_1 \neq 0$. In the solution to Exercise 10.33, we had $t = -1.66$ (df $= 22$) and $P = 0.111$, so we cannot reject H_0.

10.44. **(a)** With $r = -0.19$ and $n = 713$, we have $t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \doteq -5.16$. **(b)** Comparing to a t distribution with df $= 711$ (or anything reasonably close), the P-value is less than 0.0001, so we conclude that $\rho \neq 0$.

10.45. For linear regression, DFM $= 1$. Because DFT $=$ DFM $+$ DFE and SST $=$ SSM $+$ SSE, we can find the missing degrees of freedom (DF) and sum of squares (SS) entries on the Residual row by subtraction: DFE $= 18$ and SSE $= 3995.4$. The entries in the mean square (MS) column are MSM $= \frac{\text{SSM}}{\text{DFM}} = 4560.6$ and MSE $= \frac{\text{SSE}}{\text{DFE}} \doteq 221.97$. Finally, $F = \frac{\text{MSM}}{\text{MSE}} \doteq 20.55$.

Source	DF	SS	MS	F
Regression	1	4560.6	4560.6	20.55
Residual	18	3995.4	221.97	
Total	19	8556.0		

10.46. $s = \sqrt{\text{MSE}} \doteq 14.8985$ and $r^2 = \frac{\text{SSM}}{\text{SST}} = \frac{4560.6}{8556.0} \doteq 0.5330$.

10.47. As $s_x = \sqrt{\frac{1}{19}\sum(x_i - \bar{x})^2} = 19.99\%$, we have $\sqrt{\sum(x_i - \bar{x})^2} = s_x\sqrt{19} \doteq 87.1344\%$, so:

$$\text{SE}_{b_1} = \frac{s}{\sqrt{\sum(x_i - \bar{x})^2}} \doteq \frac{14.8985}{87.1344} \doteq 0.1710$$

Alternatively, note that we have $F \doteq 20.55$ and $b_1 = 0.775$. Because $t^2 = F$, we know that $t = 4.5332$ (take the positive square root, because t and b_1 have the same sign). Then $\text{SE}_{b_1} = b_1/t = 0.1710$. (Note that with this approach, we do not need to know that $s_x = 19.99\%$.)

Finally, with df = 18, $t^* = 2.1009$ for 95% confidence, so the 95% confidence interval is $0.775 \pm 0.3592 = 0.4158$ to 1.1342.

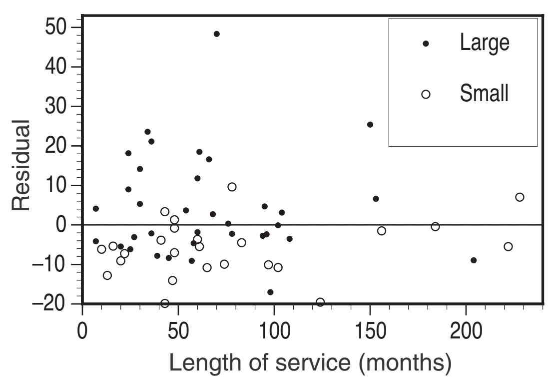
10.48. **(a)** With $\bar{x} \doteq 80.9$, $s_x \doteq 17.2$, $\bar{y} \doteq 43.5$, $s_y \doteq 20.3$, and $r \doteq 0.68$, we find:

$$b_1 = (0.68)\left(\frac{20.3}{17.2}\right) \doteq 0.8026$$
$$b_0 = 43.5 - (0.8026)(80.9) \doteq -21.4270$$

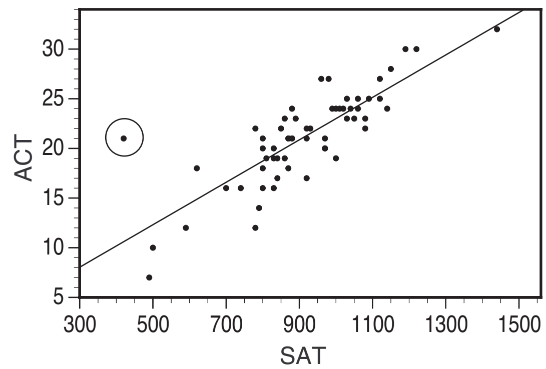
(Answers may vary slightly due to rounding.) The regression equation is therefore $\widehat{\text{GHP}} = -21.4270 + 0.8026\,\text{FVC}$. **(b)** Testing $\beta_1 = 0$ is equivalent to testing $\rho = 0$, so the test statistic is $t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \doteq 6.43$ (df = 48), for which $P < 0.0005$. The slope (correlation) is significantly different from 0.

10.49. Use the formula $t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$ with $r = 0.6$. For $n = 20$, $t = 3.18$ with df = 18, for which the two-sided P-value is $P = 0.0052$. For $n = 10$, $t = 2.12$ with df = 8, for which the two-sided P-value is $P = 0.0667$. With the larger sample size, r should be a better estimate of ρ, so we are less likely to get $r = 0.6$ unless ρ is really not 0.

10.50. Most of the small banks have negative residuals, while most large-bank residuals are positive. This means that, generally, wages at large banks are higher, and small bank wages are lower, than we would predict from the regression.



10.51. (a) Not surprisingly, there is a positive association between scores. The 47th pair of scores (circled) is an outlier—the ACT score (21) is higher than one would expect for the SAT score (420). Since this SAT score is so low, this point may be influential. No other points fall outside the pattern. **(b)** The regression equation is $\hat{y} = 1.626 + 0.02137x$. The slope is significantly different from 0: $t = 10.78$ (df = 58) for which $P < 0.0005$. **(c)** $r = 0.8167$.



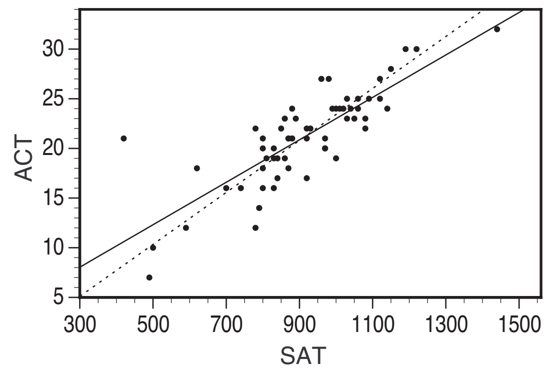
Minitab output: Regression of ACT score on SAT score

```
The regression equation is ACT = 1.63 + 0.0214 SAT

Predictor        Coef       Stdev    t-ratio        p
Constant        1.626       1.844       0.88    0.382
SAT          0.021374    0.001983      10.78    0.000

s = 2.744       R-sq = 66.7%     R-sq(adj) = 66.1%
```

10.52. (a) The means are identical (21.133). **(b)** For the observed ACT scores, $s_y = 4.714$; for the fitted values, $s_{\hat{y}} = 3.850$. **(c)** For $z = 1$, the SAT score is $\bar{x} + s_x = 912.7 + 180.1 = 1092.8$. The predicted ACT score is $\hat{y} \doteq 25$ (Minitab reports 24.983), which gives a standard score of about 1 (using the standard deviation of the *predicted* ACT scores. **(d)** For $z = -1$, the SAT score is $\bar{x} - s_x = 912.7 - 180.1 = 732.6$. The predicted ACT score is $\hat{y} \doteq 17.3$ (Minitab reports 17.285), which gives a standard score of about -1. **(e)** It appears that the standard score of the predicted value is the same as the explanatory variable's standard score. (See note below.)

Note: (a) *This will always be true because* $\sum_i \hat{y}_i = \sum_i (b_0 + b_1 x_i) = n\,b_0 + b_1 \sum_i x_i = n(\bar{y} - b_1\bar{x}) + b_1 n\,\bar{x} = n\,\bar{y}$. **(b)** *The standard deviation of the predicted values will be* $s_{\hat{y}} = |r| s_y$*; in this case,* $s_{\hat{y}} = (0.8167)(4.714)$. *To see this, observe that the variance of the predicted values is* $\frac{1}{n-1}\sum_i (\hat{y}_i - \bar{y})^2 = \frac{1}{n-1}\sum_i (b_1 x_i - b_1\bar{x})^2 = b_1^2 s_x^2 = r^2 s_y^2$. **(e)** *For a given standard score z, note that* $\hat{y} = b_0 + b_1(\bar{x} + z\,s_x) = \bar{y} - b_1\bar{x} + b_1\bar{x} + b_1 z\,s_x = \bar{y} + z\,r\,s_y$. *If* $r > 0$*, the standard score for* $\hat{y}$ *equals* z*; if* $r < 0$*, the standard score is* $-z$.

10.53. (a) For SAT: $\bar{x} = 912.\overline{6}$ and $s_x = 180.1117$. For ACT: $\bar{y} = 21.1\overline{3}$ and $s_y = 4.7137$. Therefore, the slope is $a_1 \doteq 0.02617$ and the intercept is $a_0 \doteq -2.7522$. **(b)** The new line is dashed. **(c)** For example, the first prediction is $-2.7522 + (0.02617)(1000) \doteq 23.42$. Up to rounding error, the mean and standard deviation of the predicted scores are the same as those of the ACT scores: $\bar{y} = 21.1\overline{3}$ and $s_y = 4.7137$.



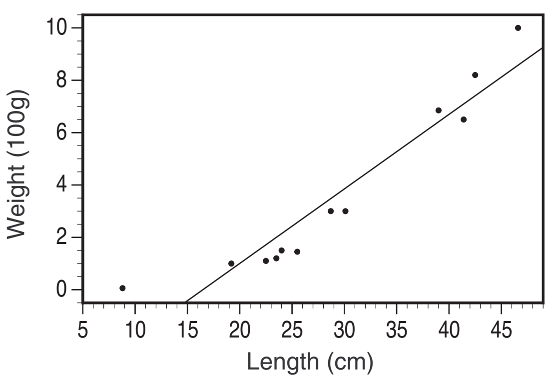
Note: *The usual least-squares line minimizes the total squared vertical distance from the points to the line. If* instead *we seek to minimize the total of* $\sum_i |h_i v_i|$, *where* h_i *is the horizontal distance and* v_i *is the vertical distance, we obtain the line* $\hat{y} = a_0 + a_1 x$—*except that we must choose the sign of* a_1 *to be the same as the sign of* r. *(It would hardly be the "best line" if we had a positive slope with a negative association.) If* $r = 0$, *either sign will do.*

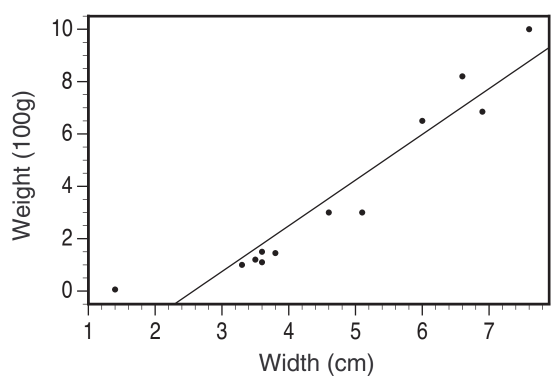
10.54. (a) The regression equations are:

$\widehat{\text{WEIGHT}} = -468.91 + 28.462\ \text{LENGTH}$ with $s \doteq 109.4$ and $r^2 \doteq 0.902$

$\widehat{\text{WEIGHT}} = -449.44 + 174.63\ \text{WIDTH}$ with $s \doteq 107.9$ and $r^2 \doteq 0.905$

(b) Both scatterplots suggest that the relationships are curved rather than linear. (Points to the left and right lie above the line; those in the middle are generally below the line.)





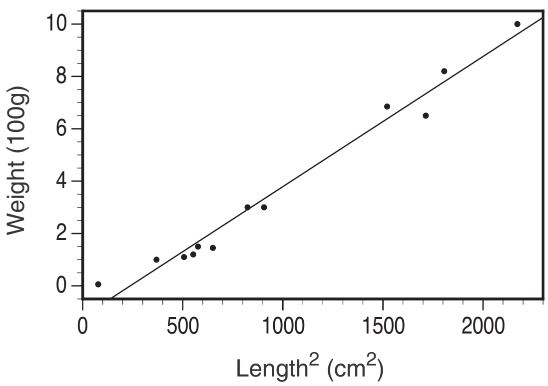
Minitab output: Regression of weight on length (Model 1)

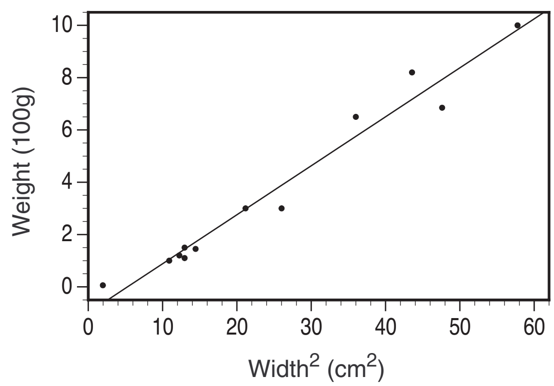
```
The regression equation is weight = -469 + 28.5 length

Predictor        Coef        Stdev    t-ratio        p
Constant      -468.91        92.55      -5.07    0.000
length         28.462        2.967       9.59    0.000

s = 109.4        R-sq = 90.2%     R-sq(adj) = 89.2%
```

Regression of weight on width (Model 2)

```
The regression equation is weight = -449 + 175 width

Predictor        Coef        Stdev    t-ratio        p
Constant      -449.44        89.27      -5.03    0.000
width          174.63        17.93       9.74    0.000

s = 107.9        R-sq = 90.5%     R-sq(adj) = 89.5%
```

10.55. **(a)** For squared length: $\widehat{\text{Weight}} = -117.99 + 0.4970\ \text{SQLEN}$, $s \doteq 52.76$, $r^2 = 0.977$.
(b) For squared width: $\widehat{\text{Weight}} = -98.99 + 18.732\ \text{SQWID}$, $s \doteq 65.24$, $r^2 = 0.965$.
Both scatterplots look more linear.





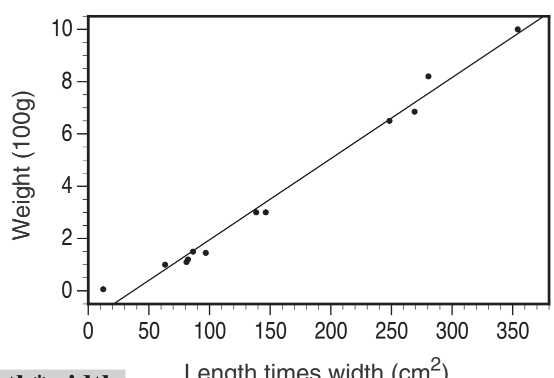
Minitab output: Regression of weight on squared length (Model 1)

```
The regression equation is weight = -118 + 0.497 sqlen

Predictor        Coef       Stdev    t-ratio        p
Constant      -117.99       27.88      -4.23    0.002
sqlen         0.49701     0.02400      20.71    0.000

s = 52.76       R-sq = 97.7%     R-sq(adj) = 97.5%
```

Regression of weight on squared width (Model 2)

```
The regression equation is weight = -99.0 + 18.7 sqwid

Predictor        Coef       Stdev    t-ratio        p
Constant       -98.99       33.67      -2.94    0.015
sqwid          18.732       1.126      16.64    0.000

s = 65.24       R-sq = 96.5%     R-sq(adj) = 96.2%
```

10.56. **(a)** The regression line is $\widehat{\text{WEIGHT}} = -115.10 + 3.1019(\text{LENGTH})(\text{WIDTH})$, $s \doteq 41.69$, $r^2 = 0.986$. **(b)** As measured by r^2, this last model is (by a slim margin) the best. (However, this scatterplot again gives some suggestion of curvature, indicating that some other model might do better still.)



Minitab output: Regression of weight on length*width

```
The regression equation is weight = -115 + 3.10 lenwid

Predictor        Coef       Stdev    t-ratio        p
Constant      -115.10       21.87      -5.26    0.000
lenwid         3.1019      0.1179      26.32    0.000

s = 41.69       R-sq = 98.6%     R-sq(adj) = 98.4%
```

10.57. The table on the right shows the correlations and the corresponding test statistics. The first two results agree with the results of (respectively) Exercises 10.22 and 10.23.

	r	t	P
IBI/area	0.4459	3.42	0.0013
IBI/forest	0.2698	1.92	0.0608
area/forest	−0.2571	−1.82	0.0745

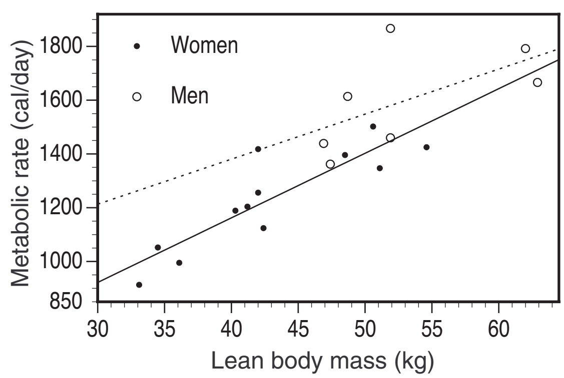
10.58. The correlation was significant for vegetables, fruit, and meat, and nearly significant for eggs. All the significant correlations are negative, meaning (for example) that children with *high* neophobia tend to eat these foods *less* frequently.

	r	t	P
Vegetables	−0.27	−6.65	0.0000
Fruit	−0.16	−3.84	0.0001
Meat	−0.15	−3.60	0.0004
Eggs	−0.08	−1.90	0.0576
Sweet/fatty snacks	0.04	0.95	0.3430
Starchy staples	−0.02	−0.47	0.6355

10.59. For each correlation, we compute $t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$. For the whole group, t ranges from 2.245 ($P = 0.0266$) to 3.208 ($P = 0.0017$). For Caucasians only, t ranges from 1.572 ($P = 0.1193$) to 2.397 ($P = 0.0185$). The three smallest correlations (0.16 and 0.19) are the only ones that are not significant.

Rule Breaking Measure	Popularity	Gene Expression
Sample 1 ($n = 123$)		
RB.composite	0.28**	0.26**
RB.questionnaire	0.22*	0.23*
RB.video	0.24**	0.20*
Sample 1 Caucasians only ($n = 96$)		
RB.composite	0.22*	0.23*
RB.questionnaire	0.16	0.24*
RB.video	0.19	0.16

10.60. See also the solution to Exercise 2.35. **(a)** The association is linear and positive; the women's points show a stronger association. As a group, males typically have larger values for both variables. **(b)** The women's regression line (the solid line in the graph) is $\hat{y} = 201.2 + 24.026x$, with $s \doteq 95.08$ and $r^2 = 0.768$. The men's line (the dashed line) is $\hat{y} = 710.5 + 16.75x$, with $s \doteq 167.1$ and $r^2 = 0.351$. The women's slope is significantly different from 0 ($t = 5.76$, df = 10, $P < 0.0005$), but the men's is not ($t = 1.64$, df = 5, $P = 0.161$). These test results, and the values of s and r^2, confirm the observation that the women's association is stronger—however, see the solution to the next exercise.



10.61. **(a)** These intervals (in the table below) overlap quite a bit. **(b)** These quantities can be computed from the data, but it is somewhat simpler to recall that they can be found from the sample standard deviations $s_{x,\text{w}}$ and $s_{x,\text{m}}$:

$$s_{x,\text{w}}\sqrt{11} \doteq 6.8684\sqrt{11} \doteq 22.78 \quad \text{and} \quad s_{x,\text{m}}\sqrt{6} \doteq 6.6885\sqrt{6} \doteq 16.38$$

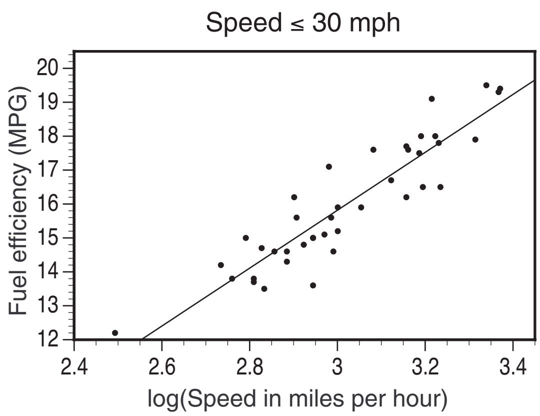
The women's SE_{b_1} is smaller in part because it is divided by a large number. **(c)** In order to reduce SE_{b_1} for men, we should choose our new sample to include men with a wider variety of lean body masses. (Note that just taking a larger sample will reduce SE_{b_1}; it is reduced even *more* if we choose subjects who will increase $s_{x,\text{m}}$.)

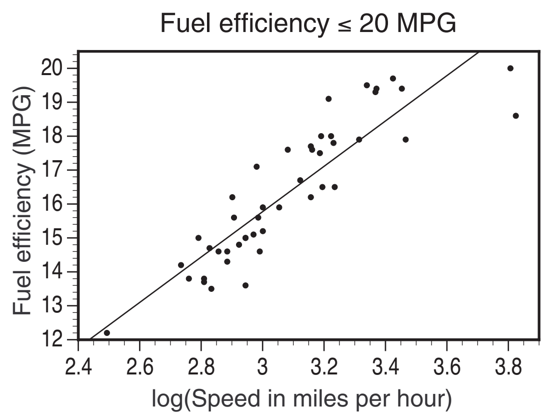
	b_1	SE_{b_1}	df	t^*	Interval
Women	24.026	4.174	10	2.2281	14.7257 to 33.3263
Men	16.75	10.20	5	2.5706	−9.4699 to 42.9699

10.62. Scatterplots, and portions of the Minitab outputs, are shown below. The equations are:

For all points,	$\widehat{\text{MPG}} = -7.796 + 7.8742\ \text{LOGMPH}$
For speed ≤ 30 mph,	$\widehat{\text{MPG}} = -9.786 + 8.5343\ \text{LOGMPH}$
For fuel efficiency ≤ 20 mpg,	$\widehat{\text{MPG}} = -4.282 + 6.6854\ \text{LOGMPH}$

Students might make a number of observations about the effects of the restrictions; for example, the estimated coefficients (and their standard errors) change quite a bit.





Minitab output: Regression of fuel efficiency on log(speed)— with all points

```
Predictor        Coef        Stdev    t-ratio        p
Constant       -7.796        1.155      -6.75    0.000
logMPH         7.8742       0.3541      22.24    0.000

s = 0.9995      R-sq = 89.5%     R-sq(adj) = 89.3%
```

. . . with speed 30 mph or less

```
Predictor        Coef        Stdev    t-ratio        p
Constant       -9.786        1.862      -5.26    0.000
logMPH         8.5343       0.6154      13.87    0.000

s = 0.7600      R-sq = 83.5%     R-sq(adj) = 83.1%
```

. . . with fuel efficiency 20 mpg or less

```
Predictor        Coef        Stdev    t-ratio        p
Constant       -4.282        1.647      -2.60    0.013
logMPH         6.6854       0.5323      12.56    0.000

s = 0.9462      R-sq = 78.6%     R-sq(adj) = 78.1%
```

Chapter 11 Solutions

11.1. **(a)** The response variable is math GPA. **(b)** The number of cases is $n = 106$. **(c)** There were $p = 4$ explanatory variables. **(d)** The explanatory variables were SAT Math, SAT Verbal, class rank, and mathematics placement score.

11.2. **(a)** $\hat{y} = -3.8 + 7.3(3) - 2.1(1) = 16$. **(b)** No: We can compute predicted values for any values of x_1 and x_2. (Of course, it helps if they are close to those in the data set.) **(c)** This is determined by the coefficient of x_1: An increase of two units in x_1 results in an increase of $(7.3)(2) = 14.6$ units in $\hat{y}$.

11.3. **(a)** The fact that the coefficients are all positive indicates that math GPA should increase when any explanatory variable increases (as we would expect). **(b)** With $n = 86$ cases and $p = 4$ variables, DFM $= p = 4$ and DFE $= n - p - 1 = 81$. **(c)** In the following table, each t statistic is the estimate divided by the standard error; the P-values are computed from a t distribution with df $= 81$. (The t statistic for the intercept was not required for this exercise, but is included for completeness.)

Variable	Estimate	SE	t	P
Intercept	−0.764	0.651	−1.1736	0.2440
SAT Math	0.00156	0.00074	2.1081	0.0381
SAT Verbal	0.00164	0.00076	2.1579	0.0339
HS rank	1.470	0.430	3.4186	0.0010
Bryant placement	0.889	0.402	2.2114	0.0298

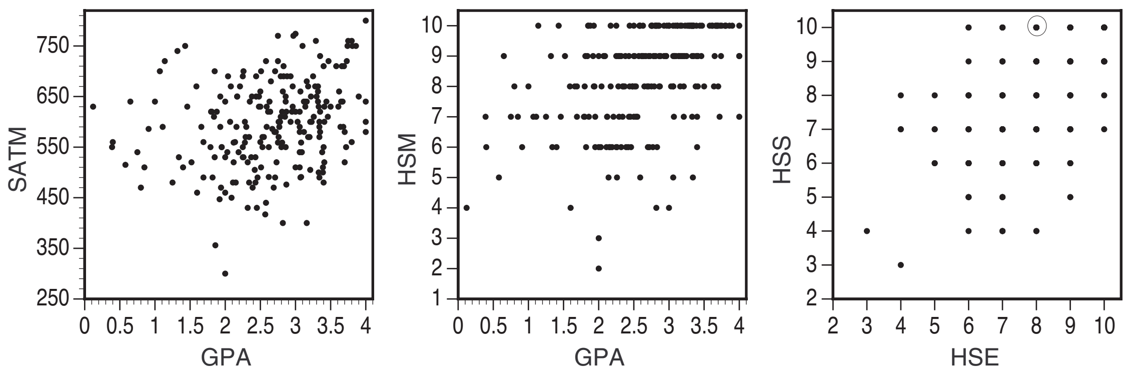
All four coefficients are significantly different from 0 (although the intercept is not).

11.4. The missing entries in the DF and SS columns can be found by noting that DFE + DFM = DFT and SSE + SSM = SST. The MS (mean square) entries are computed as SS divided by DF, and $F = \text{MSM}/\text{MSE}$. Comparison of $F = 2.84$ to an F distribution with df 3 and 50 gives $P \doteq 0.0471$, so we conclude the regression is significant at the 5% level. Finally, $R^2 = \frac{\text{SSM}}{\text{SST}} = \frac{75}{515} \doteq 0.1456$.

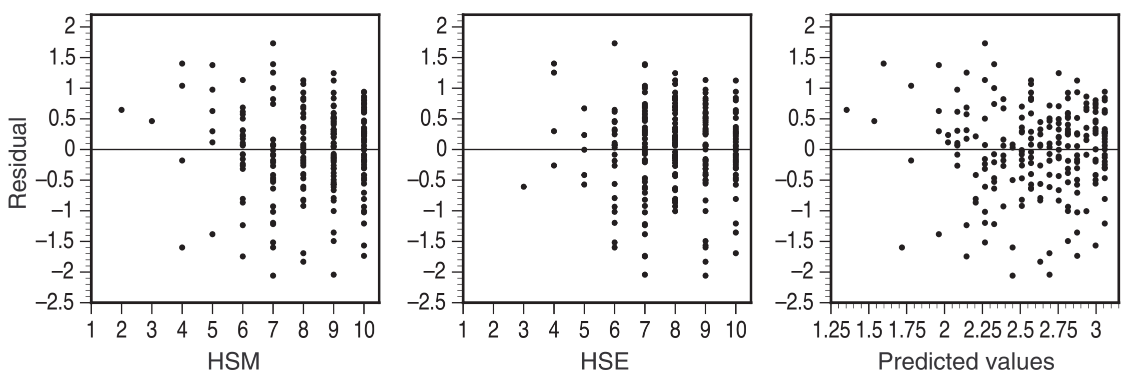
Source	DF	SS	MS	F
Model	3	75	25	2.84
Error	50	440	8.8	
Total	53	515		

11.5. The correlations are found in Figure 11.4 and are summarized in the table on the right. Of the 15 possible scatterplots to be made from these six variables, three are shown below as examples. The pairs with the largest correlations are generally easy to pick out. The whole-number scale for high school grades causes point clusters in those scatterplots and makes it difficult to determine the strength of the association. For example, in the plot of HSS versus HSE below, the circled point represents 9 of the 224 students. One might guess that these three scatterplots show relationships of roughly equal strength, but because of the overlapping points, the correlations are quite different; from left to right, they are 0.2517, 0.4365, and 0.5794.

	SATM	SATV	HSM	HSS	HSE
GPA	0.2517	0.1145	0.4365	0.3294	0.2890
SATM		0.4639	0.4535	0.2405	0.1083
SATV			0.2211	0.2617	0.2437
HSM				0.5757	0.4469
HSS					0.5794



11.6. The regression equation is given in the Minitab output below. The whole-number scale for high school grades means that the predicted values also come in clusters. All but 21 students had both HSM and HSE above 5, so for all three plots, there are few residuals on the left half.



Minitab output: Regression of GPA on HSM and HSE

```
The regression equation is GPA = 0.624 + 0.183 HSM + 0.0607 HSE

Predictor        Coef       Stdev    t-ratio        p
Constant       0.6242      0.2917       2.14    0.033
HSM           0.18265     0.03196       5.72    0.000
HSE           0.06067     0.03473       1.75    0.082

s = 0.6996      R-sq = 20.2%     R-sq(adj) = 19.4%
```

11.7. The table below gives two sets of answers: those found with critical values from Table D and those found with software. In each case, the estimated coefficient is $b_1 = 6.4$ with standard error $SE_{b_1} = 3.1$, and the margin of error is $t^* SE_{b_1}$, with df $= n - 3$ for parts (a) and (b), and df $= n - 4$ for parts (c) and (d). (The Table D interval for part (d) uses df $= 100$.)

			Table D		Software	
	n	df	t^*	Interval	t^*	Interval
(a)	27	24	2.064	0.0016 to 12.7984	2.0639	0.0019 to 12.7981
(b)	53	50	2.009	0.1721 to 12.6279	2.0086	0.1735 to 12.6265
(c)	27	23	2.069	−0.0139 to 12.8139	2.0687	−0.0128 to 12.8128
(d)	124	120	1.984	0.2496 to 12.5504	1.9799	0.2622 to 12.5378

11.8. For all four settings, the test statistic is $t = b_1/SE_{b_1} = 6.4/3.1 \doteq 2.065$, with df $= n - 3$ for parts (a) and (b) and df $= n - 4$ for parts (c) and (d). The P-values are 0.0499, 0.0442, 0.0504, and 0.0411. At the 5% significance level, we would reject the null hypothesis for each test except (c); the test is barely significant for (a), and barely not significant for (c). (This is consistent with the confidence intervals from the previous exercise.)

11.9. **(a)** H_0 should refer to β_2 (the population coefficient) rather than b_2 (the estimated coefficient). **(b)** This sentence should refer to the *squared* multiple correlation. **(c)** A small P implies that *at least one coefficient* is different from 0.

11.10. **(a)** Multiple regression only assumes Normality of the error terms (residuals), not the explanatory variables. (The explanatory variables do not even need to be random variables.) **(b)** A small P-value tells us that the model is significant (useful for prediction) but does not measure its explanatory power (the accuracy of those predictions). The squared multiple correlation R^2 is a measure of explanatory power. **(c)** For example, if x_1 and x_2 are significantly correlated with each other and with the response variable, it might turn out that the coefficient of x_1 is statistically significant and the coefficient of x_2 is not. **(d)** R is not the average correlation; if it were, adding additional variables might reduce make R closer to 0. R^2 tells us the total explanatory power of the entire model.

Note: *The statement for part (c) is a paraphrase of the "Caution" on page 602 of the text. As a simple illustration of how this might happen, suppose that the response variable* $y = ax_1 + b$ *(with little or no error term), where all observed values of* x_1 *are positive, and the second explanatory variable is* $x_2 = x_1^2$*. The correlation between y and* x_2 *might be very large, but in a multiple regression model with* x_1*, the coefficient of* x_2 *will almost certainly not be significant.*

11.11. **(a)** $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_8 x_{i8} + \epsilon_i$, where $i = 1, 2, \ldots, 135$, and ϵ_i are independent $N(0, \sigma)$ random variables. **(b)** The sources of variation are model (DFM $= p = 8$), error (DFE $= n - p - 1 = 126$), and total (DFT $= n - 1 = 134$).

11.12. **(a)** With $n = 82$ and $p = 6$, the degrees of freedom in the ANOVA table are DFM $= p = 6$, DFE $= n - p - 1 = 75$, and DFT $= n - 1 = 81$. With the first two degrees of freedom, we can find MSM $= \frac{\text{SSM}}{\text{DFM}} \doteq 3.7667$ and MSE $= \frac{\text{SSE}}{\text{DFE}} = 1.34$, and then compute $F = \frac{\text{MSM}}{\text{MSE}} \doteq 2.81$. **(b)** This F statistic has df 6 and 75. **(c)** Comparing to the $F(6, 60)$ critical values in Table E, we note that $2.63 < F < 3.12$, so $0.01 < P < 0.025$. (Software gives 0.016.) **(d)** This regression explains $R^2 = \frac{\text{SSM}}{\text{SST}} \doteq 18.4\%$ of the variation in the response variable.

Source	DF	SS	MS	F
Model	6	22.6	3.7667	2.8109
Error	75	100.5	1.3400	
Total	81	123.1		

11.13. We have $p = 8$ explanatory variables and $n = 795$ observations. **(a)** The ANOVA F test has degrees of freedom DFM $= p = 8$ and DFE $= n - p - 1 = 786$. **(b)** This model explains only $R^2 \doteq 7.84\%$ of the variation in energy-drink consumption; it is not very predictive. **(c)** A positive (negative) coefficient means that large values of that variable correspond to higher (lower) energy-drink consumption. Therefore, males and Hispanics consume energy drinks more frequently, and consumption increases with risk-taking scores. **(d)** Within a group of students with identical (or similar) values of those other variables, energy-drink consumption increases with increasing jock identity and increasing risk taking.

11.14. No (or at least, not necessarily). It is possible that, although no individual coefficient is significant, the whole group (or some subset) is. Recall that the t tests "assess the significance of each predictor variable assuming that all other predictors are included in the regression equation." If one variable is removed from the model (because its t statistic is not significant), we can no longer use the other t statistics to draw conclusions about the remaining coefficients.

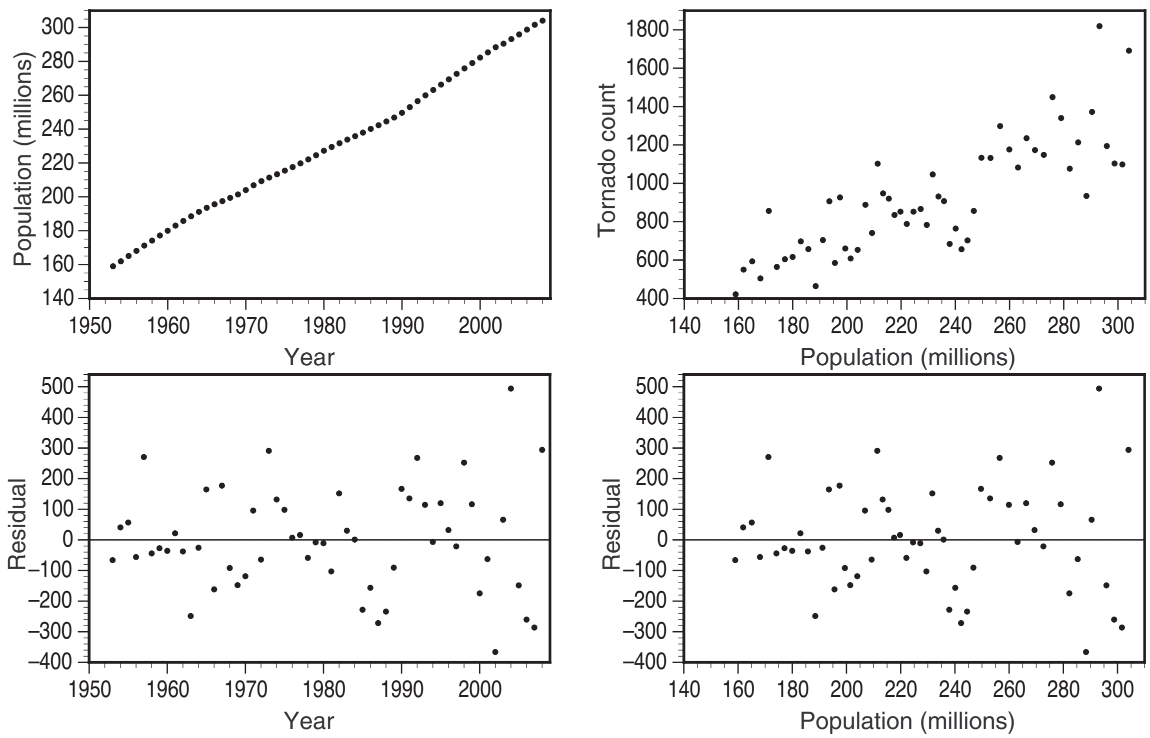
11.15. We have $n = 202$, and $p = 1$ (for Model 1) or $p = 2$ (for Model 2). **(a)** For Model 1, DFE $= 200$. For Model 2, DFE $= 199$. **(b)** and **(c)** The test statistics $t = b_i/\text{SE}_{b_i}$ and P-values are in the table on the right. **(d)** The relationship is still positive after adjusting for RB. When gene expression increases by 1, popularity increases by 0.204 in Model 1, and by 0.161 in Model 2 (with RB fixed).

Model	Variable	t	P
1	Gene expression	$\frac{0.204}{0.066} \doteq 3.09$	0.0023
2	Gene expression	$\frac{0.161}{0.066} \doteq 2.44$	0.0153
	RB	$\frac{0.100}{0.030} \doteq 3.33$	0.0010

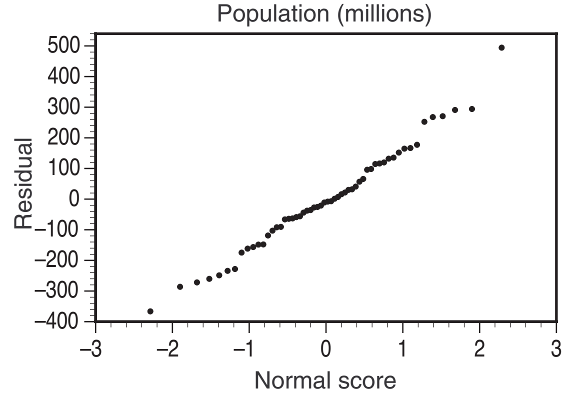
11.16. **(a)** All three correlations quite high: year and tornado count (0.8095), population and tornado count (0.8180), and year and population (0.9981). The solution to Exercise 10.19 shows a scatterplot of tornadoes versus year; the other two scatterplots are shown on the following page. Because of the near-perfect linear relationship between year and population, the plot of tornadoes versus population looks nearly identical to the plot of tornadoes versus year (apart from horizontal scale). **(b)** The regression equation is $\widehat{\text{COUNT}} = 63677 - 33.91\ \text{YEAR} + 0.0191\ \text{CENSUS}$. (Minitab output on the following page.) **(c)** The only cause for concern in the analysis is the extremely high count from 2004, which is visible in all the plots. The plots versus year and versus population are nearly identical, apart from scale; neither plot shows any striking patterns. The Normal quantile plot (along with a stemplot of the residuals) suggests no serious deviations from Normality. **(d)** To look for a linear increase over time, we test H_0: $\beta_1 = 0$ versus H_a: $\beta_1 \neq 0$, where β_1 is coefficient of YEAR in our model. The test statistic is $t = -1.46$ ($P = 0.149$), so we cannot reject H_0. With population included, the predictive information in year is made redundant. (That is, once we know the population, the additional information from year does not appreciably improve our estimate of tornado count.)

Minitab output: Regression of tornadoes on year and population

```
Count = 63677 - 33.9 Year + 0.0191 Census

Predictor        Coef       Stdev    t-ratio        p
Constant        63677       43769       1.45    0.152
Year           -33.91       23.15      -1.46    0.149
Census       0.019124    0.009068       2.11    0.040

s = 171.5       R-sq = 68.2%     R-sq(adj) = 67.0%
```



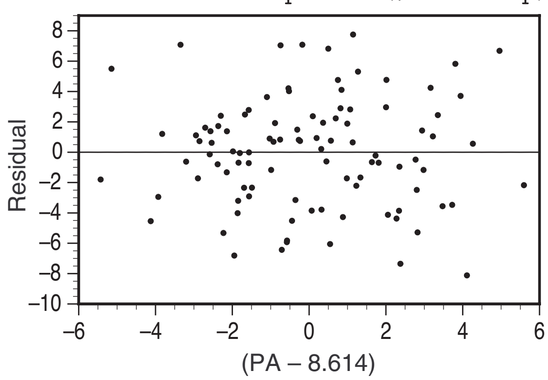
−3	6
−2	876432
−1	7654410
−0	9966655433222100
0	00123345699
1	111335667
2	56799
3	
4	9

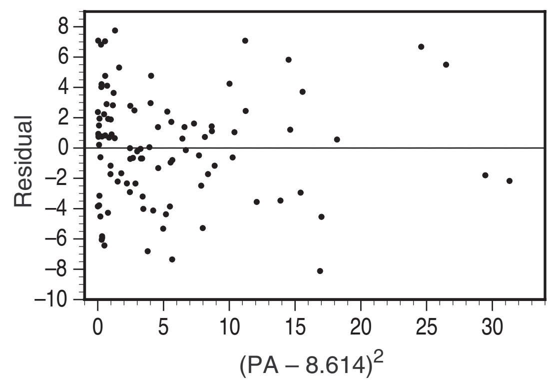


11.17. **(a)** The regression equation is $\widehat{\text{BMI}} = 23.4 - 0.682x_1 + 0.102x_2$. (Minitab output below.) **(b)** The quadratic regression explains $R^2 \doteq 17.7\%$ of the variation in BMI. **(c)** Analysis of residuals might include a stemplot, plots of residuals versus x_1 and x_2, and a Normal quantile plot. All of these appear below; none suggest any obvious causes for concern. **(d)** From the Minitab output, $t = 1.83$ with df $= 97$, for which $P = 0.070$—not significant.

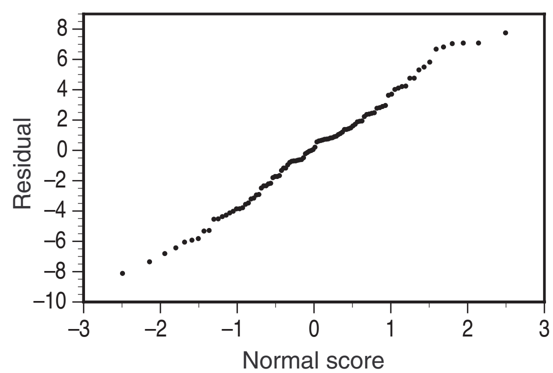
Minitab output: Quadratic regression for predicting BMI from PA

```
The regression equation is BMI = 23.4 - 0.682 X1 + 0.102 X2

Predictor        Coef        Stdev    t-ratio        p
Constant      23.3956       0.4670      50.10    0.000
X1            -0.6818       0.1572      -4.34    0.000
X2            0.10195      0.05556       1.83    0.070

s = 3.612        R-sq = 17.7%     R-sq(adj) = 16.0%
```





−8	1
−7	3
−6	840
−5	9832
−4	553210
−3	8875421
−2	9943321
−1	7776311
−0	9776666642100
0	0256667778889
1	012334467899
2	233447889
3	67
4	012277
5	348
6	68
7	0007

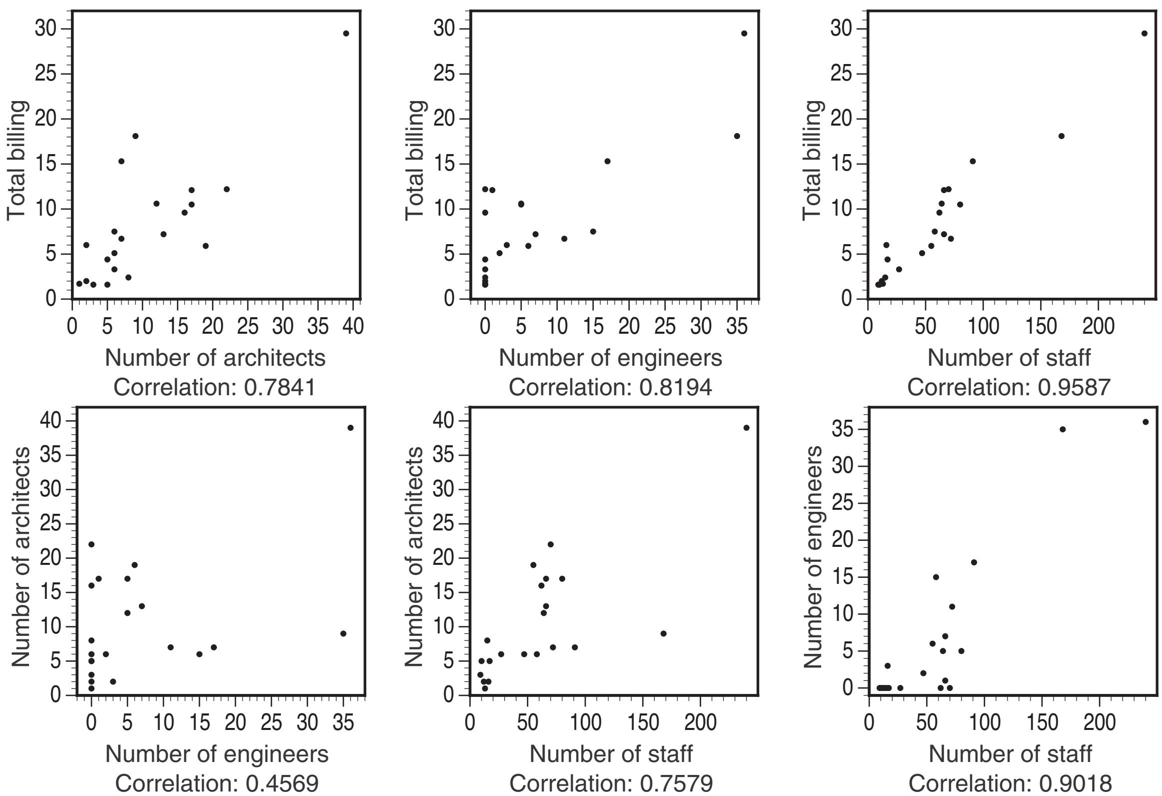


11.18. **(a)** All distributions are skewed to the right (stemplots follow). Student choices of summary statistics may vary; five-number summaries are a good choice because of the skewness, but some may also give means and standard deviations. Notice especially how the skewness is apparent in the five-number summaries.

Variable	$\bar{x}$	s	Min	Q_1	M	Q_3	Max
Total billing	8.2524	6.7441	1.6	2.85	6.7	11.35	29.5
Number of architects	10.5714	8.9026	1.0	5.00	7.0	16.50	39.0
Number of engineers	6.8095	10.7964	0.0	0.00	2.0	9.00	36.0
Number of staff	59.9048	55.8891	9.0	15.50	58.0	71.00	240.0

(b) Correlation coefficients are given with the scatterplots (below). All pairs of variables are positively correlated. **(c)** The regression equation is $\widehat{\text{Billing}} = 0.7799 + 0.0143\ \text{Arch} - 0.1364\ \text{Eng} + 0.1377\ \text{Staff}$, and the standard error is $s \doteq 1.935$. (Minitab output below.) **(d)** The plots of residuals versus the explanatory variables (not shown) reveal no particular causes for concern. A stemplot of the residuals (below) is somewhat right-skewed; this can also be seen in a Normal quantile plot (not shown). **(e)** The predicted billing for HCO is 3.028 (million dollars).

```
Total billing       Architects       Engineers          Staff             Residuals
0 | 1112234         0 | 1223         0 | 000000000123   0 | 011111124     -2 | 5
0 | 5566779         0 | 556667789    0 | 5567           0 | 5566667789    -1 | 99921
1 | 0022            1 | 23           1 | 1              1 |               -0 | 886544
1 | 58              1 | 6779         1 | 57             1 | 6              0 | 006
2 |                 2 | 2            2 |                2 | 4              1 | 245
2 | 9               2 |              2 |                                   2 | 1
                    3 |              3 |                                   3 | 3
                    3 | 9            3 | 56                                4 | 2
```



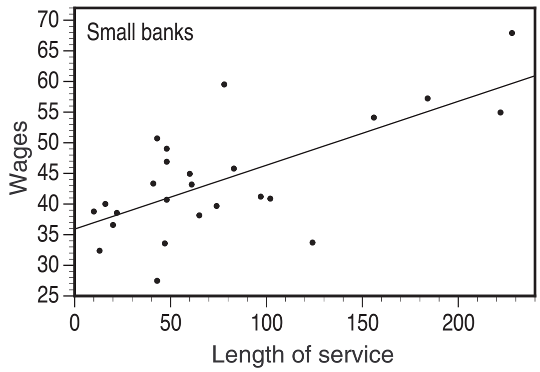
Minitab output: Regression of total billing on numbers of architects, engineers, and staff

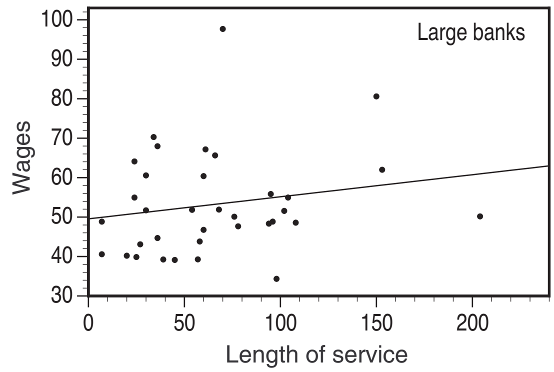
```
TotalBil = 0.780 + 0.014 N_Arch - 0.136 N_Eng + 0.138 N_Staff

Predictor        Coef       Stdev    t-ratio        p
Constant       0.7799      0.7126       1.09    0.289
N_Arch         0.0143      0.1252       0.11    0.910
N_Eng         -0.1364      0.1558      -0.88    0.393
N_Staff       0.13773     0.04104       3.36    0.004

s = 1.935       R-sq = 93.0%     R-sq(adj) = 91.8%
```

Prediction for 3 architects, 1 engineer, 17 staff members

```
     Fit  Stdev.Fit        95.0% C.I.           95.0% P.I.
   3.028      0.566    (   1.833,   4.223)  (  -1.227,   7.283)
```

11.19. **(a)** In the two scatterplots (below), we see a moderate positive linear relationship for small banks. For large banks, the relationship is very weak. **(b)** For small banks, $\widehat{\text{Wages}} = 35.9 + 0.1042\,\text{LOS}$, with $R^2 \doteq 46.6\%$ and $s \doteq 7.026$. **(c)** For large banks, $\widehat{\text{Wages}} = 49.5 + 0.0560\,\text{LOS}$, with $R^2 \doteq 3.5\%$ and $s \doteq 13.02$. **(d)** The large-bank regression is not significant (nor is it useful for prediction).





Minitab output: Regression of wages on length of service (small banks)

```
The regression equation is Wages-S = 35.9 + 0.104 LOS-S

Predictor        Coef        Stdev    t-ratio        p
Constant       35.914        2.284      15.73    0.000
LOS-S         0.10424      0.02328       4.48    0.000

s = 7.026       R-sq = 46.6%     R-sq(adj) = 44.3%
```

Regression of wages on length of service (large banks)

```
The regression equation is Wages-L = 49.5 + 0.0560 LOS-L

Predictor        Coef        Stdev    t-ratio        p
Constant       49.545        4.013      12.35    0.000
LOS-L         0.05595      0.05116       1.09    0.282

s = 13.02       R-sq = 3.5%      R-sq(adj) = 0.6%
```

11.20. **(a)** Note that most statistical software provides a way to define new variables like this. **(b)** The regression equation is $\widehat{\text{Wages}} = 35.9 + 0.1042\,\text{LOS} + 13.6\,\text{SIZE1} - 0.0483\,\text{LOSSIZE1}$. **(c)** The intercept and coefficient of LOS in this equation are the same as in the small-bank regression from the previous exercise. **(d)** Up to rounding error, these two sums equal the intercept and coefficient of LOS in the large-bank regression: Adding the intercept and SIZE1 coefficient gives $35.9 + 13.6 = 49.5$, and adding the LOS and LOSSIZE1 coefficients gives $0.1042 - 0.0483 = 0.0559$. **(e)** Large banks (SIZE1 = 1) have a larger intercept, suggesting that on the average, they offer a higher starting wages (for employees with LOS = 0). However, they also have a smaller slope, meaning that (on the average) wages increase faster at smaller banks.

Minitab output: Regression of wages on LOS, SIZE1, and LOSSIZE1

```
Wages = 35.9 + 0.104 LOS + 13.6 size1 - 0.0483 LOSsize1

Predictor        Coef        Stdev    t-ratio        p
Constant       35.914        3.562      10.08    0.000
LOS           0.10424      0.03632       2.87    0.006
SIZE1          13.631        4.910       2.78    0.007
LOSSIZE1     -0.04828      0.05634      -0.86    0.395

s = 10.96       R-sq = 26.6%     R-sq(adj) = 22.7%
```

11.21. (a) Budget and Opening are right-skewed; Theaters and Opinion are roughly symmetric (slightly left-skewed). Five-number summaries are best for skewed distributions, but all possible numerical summaries are given here.

Variable	$\bar{x}$	s	Min	Q_1	M	Q_3	Max
Budget	61.81	52.47	6.5	20.0	45.0	85.0	185.0
Opening	28.59	31.89	1.1	10.0	18.6	32.1	158.4
Theaters	2785	921	808.0	2123.0	2808.0	3510.0	4366.0
Opinion	6.440	1.064	3.6	5.9	6.6	7.0	8.9

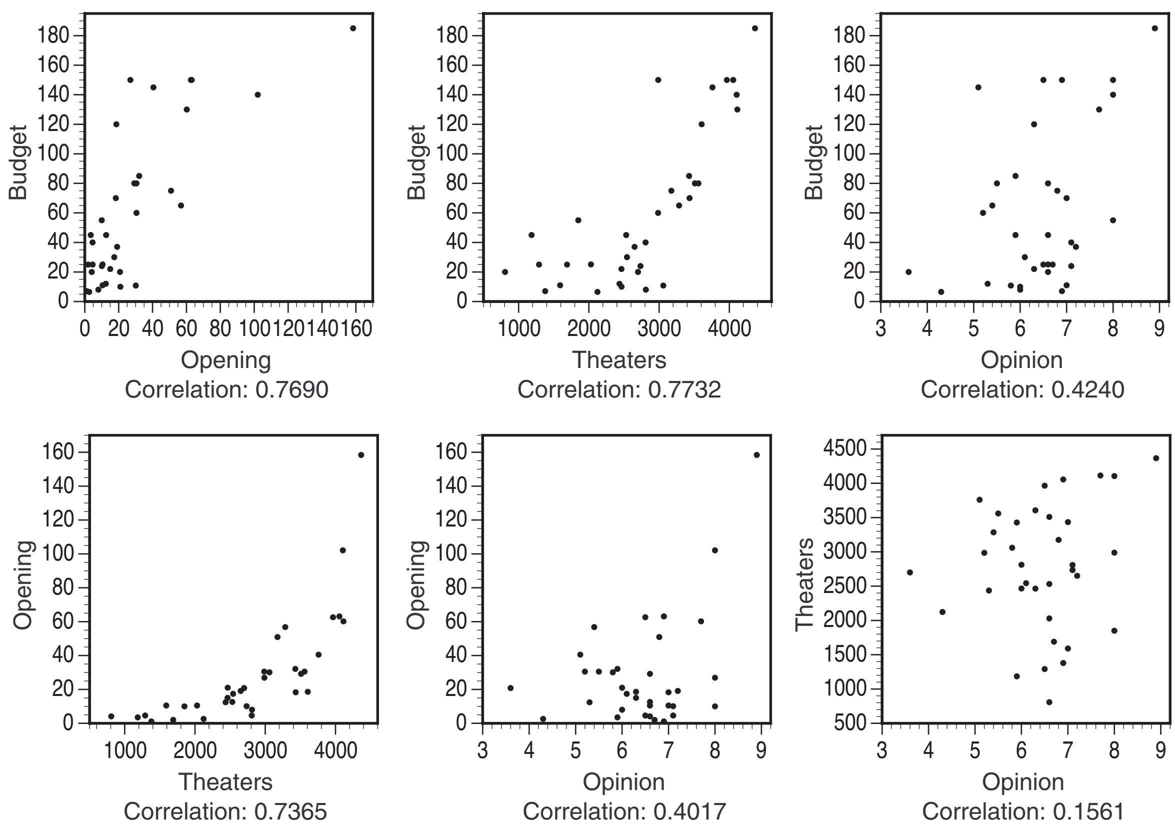
A worthwhile observation is that for all four variables, the maximum observation comes from *The Dark Knight*. **(b)** Correlation coefficients are given with the scatterplots (below). All pairs of variables are positively correlated. The Budget/Theaters and Opening/Theaters relationships appear to be curved; the others are reasonably linear.

```
Budget
0 | 0001111
0 | 222222233
0 | 4445
0 | 6677
0 | 888
1 |
1 | 23
1 | 44555
1 |
1 | 8
```

```
Opening
0 | 00000000011111111111
0 | 22223333
0 | 455
0 | 666
0 |
1 | 0
1 |
1 | 5
```

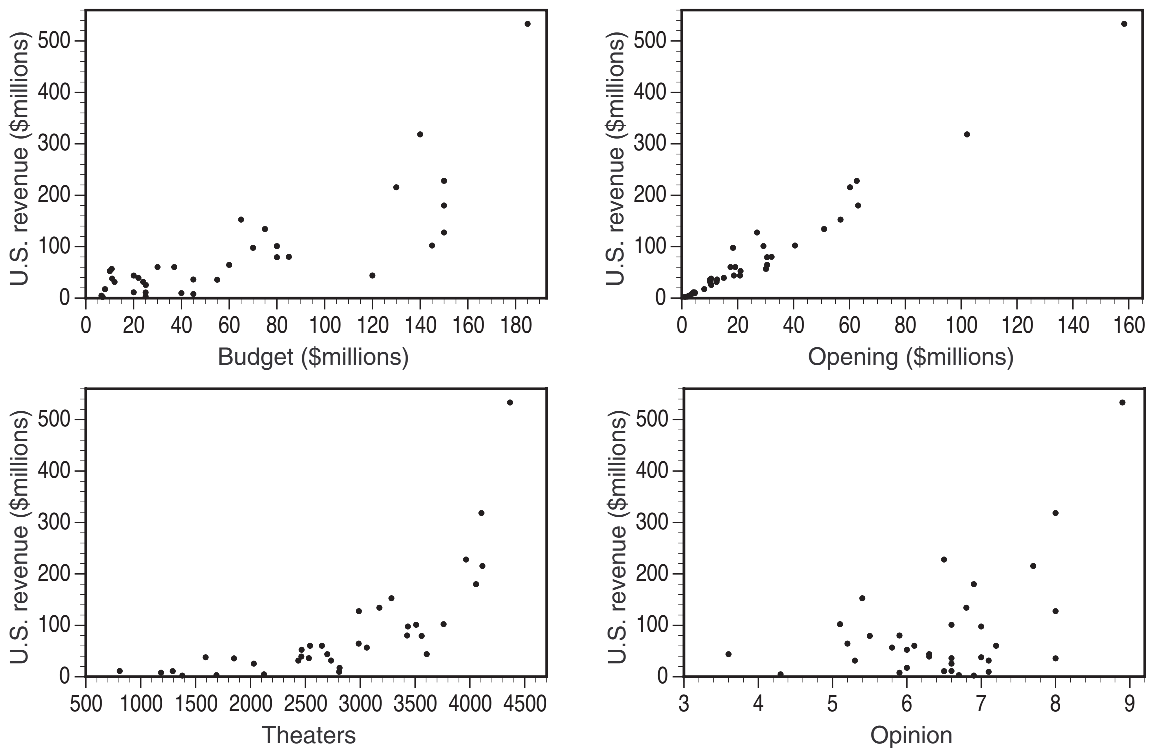
```
Theaters
0 | 8
1 | 123
1 | 568
2 | 01444
2 | 556778899
3 | 01244
3 | 55679
4 | 0113
```

```
Opinion
3 | 6
4 | 3
4 |
5 | 1234
5 | 5899
6 | 00133
6 | 5566667899
7 | 00112
7 | 7
8 | 000
8 | 9
```



11.22. **(a)** The distribution is sharply right-skewed. In millions of dollars, the mean and standard deviation are $\bar{x} \doteq 86.8$ and $s \doteq 106.2$, and the five-number summary is Min $= 2.3$, $Q_1 = 25.5$, $M = 52.5$, $Q_3 = 102.2$, Max $= 533.3$. (As in the previous exercise, the maximum revenue is for *The Dark Knight*.) **(b)** It is the deviations (errors) that need to be Normally distributed, not the response variable. **(c)** All four scatterplots (below) suggest positive associations, but only one (revenue versus opening) looks convincingly linear. Revenue versus theaters appears to be curved, and the other two are indeterminate.

0	000001112333333344
0	55666789
1	0023
1	58
2	12
2	
3	1
3	
4	
4	
5	3



11.23. **(a)** The model is

$$\text{USRevenue}_i = \beta_0 + \beta_1 \text{ Budget}_i + \beta_2 \text{ Opening}_i + \beta_3 \text{ Theaters}_i + \beta_4 \text{ Opinion}_i + \epsilon_i$$

where $i = 1, 2, \ldots, 35$, and ϵ_i are independent $N(0, \sigma)$ random variables. **(b)** The regression equation is

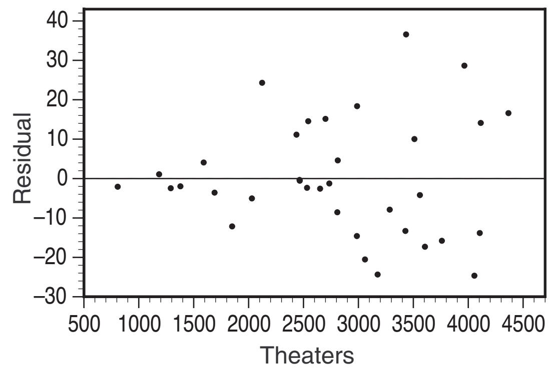
$$\widehat{\text{USRevenue}} = -67.72 + 0.1351 \text{ Budget} + 3.0165 \text{ Opening} - 0.00223 \text{ Theaters} + 10.262 \text{ Opinion}$$

(Minitab output on the following page.) **(c)** On the following page is a stemplot of the residuals. The distribution is somewhat irregular, but a Normal quantile plot (not shown) does not suggest severe deviations from Normality. The residual analysis should also include a plot of residuals versus the explanatory variables; three of those plots are unremarkable (and not shown). The plot of residuals versus theaters suggests that the spread of the residuals increases with Theaters. *The Dark Knight*—noted as unusual in the previous two exercises—may be influential. **(d)** This regression explains $R^2 \doteq 98.1\%$ of the variation in revenue.

Minitab output: Regression of U.S. Revenue on budget, opening, theaters, and opinion

```
USRevenue = - 67.7 + 0.135 Budget + 3.02 Opening - 0.00223 Theaters
           + 10.3 Opinion

Predictor       Coef       Stdev    t-ratio       p
Constant      -67.72       24.14      -2.81   0.009
Budget       0.13511     0.09776       1.38   0.177
Opening       3.0165      0.1461      20.65   0.000
Theaters   -0.002229    0.005299      -0.42   0.677
Opinion       10.262       3.032       3.38   0.002

s = 15.69       R-sq = 98.1%     R-sq(adj) = 97.8%
```

```
–2 | 440
–1 | 75
–1 | 4332
–0 | 875
–0 | 4322221100
 0 | 144
 0 |
 1 | 0144
 1 | 568
 2 | 4
 2 | 8
 3 |
 3 | 6
```



11.24. (a) The regression equation is

$$\widehat{\text{USRevenue}} = -76.6 + 3.12\,\text{Opening} + 11.5\,\text{Opinion}$$

(b) This regression explains $R^2 \doteq 97.9\%$ of the variation in U.S. revenue. **(c)** With $n = 35$, $p = 4$, $q = 2$, $R_1^2 \doteq 0.981$, and $R_2^2 \doteq 0.979$, we have

$$F = \left(\frac{n-p-1}{q}\right)\left(\frac{R_1^2 - R_2^2}{1 - R_1^2}\right) = \frac{30}{2} \cdot \frac{0.981 - 0.979}{1 - 0.981} \doteq 1.579$$

Comparing to an $F(2, 30)$ distribution, we have $P \doteq 0.2229$. We do not reject H_0, meaning that the variables removed from the regression do not add "significant predictive information" to the model.

Note: *The first printing of the text mistakenly said n = 40 movies; this changes F to 1.842 and P to 0.1735, but the conclusion is the same.*

Minitab output: Regression of U.S. revenue on opening and opinion

```
USRevenue = - 76.6 + 3.12 Opening + 11.5 Opinion

Predictor       Coef       Stdev    t-ratio       p
Constant      -76.60       17.12      -4.47   0.000
Opening      3.12355     0.09224      33.86   0.000
Opinion       11.497       2.764       4.16   0.000

s = 15.71       R-sq = 97.9%     R-sq(adj) = 97.8%
```

11.25. **(a)** Using the full model, the 95% prediction interval is \$86.86 to \$154.91 million. **(b)** With the reduced model, the interval is \$89.93 to \$155.00 million. **(c)** The intervals are very similar; as we saw in the previous exercise, there is little additional predictive information from the two variables we removed.

Note: *According to* `http://www.imdb.com/title/tt0425061/business`, *the actual U.S. revenue for* Get Smart *was \$130.3 million.*

Minitab output: Predicting U.S. revenue for *Get Smart* (full model)

```
    Fit  Stdev.Fit       95.0% C.I.           95.0% P.I.
 120.89       5.58   (  109.48,  132.29)  (   86.86,  154.91)
```

Predicting U.S. revenue for *Get Smart* (reduced model)

```
    Fit  Stdev.Fit       95.0% C.I.           95.0% P.I.
 122.46       2.86   (  116.64,  128.29)  (   89.93,  155.00)
```

11.26. **(a)** The two films are *Yes Man* and *Hancock,* which made (respectively) \$36.7 and \$34.2 million more than predicted. (The easiest way to find these two movies is to find the two largest residuals of the reduced-model regression.) **(b)** With those movies removed, the regression equation is

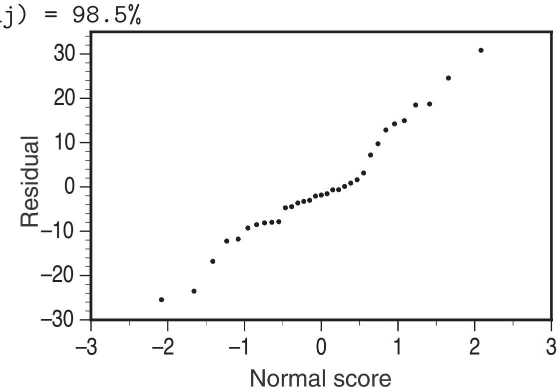
$$\widehat{\text{USRevenue}} = -75.72 + 3.1038\ \text{Opening} + 11.112\ \text{Opinion}$$

The coefficients are significant ($t = 39.15$ and $t = 4.75$, both with $P < 0.0005$). **(c)** Both coefficients decreased slightly, meaning that a change in either variable makes a slightly smaller change on the predicted revenue. Another observation: R^2 is slightly larger for this regression (98.6% versus 97.9%). This does not mean that this regression is better; rather, removing the outliers means that there is less variation to explain. **(d)** A stemplot and quantile plot (below) suggest no reasons for concern.

Minitab output: Regression of U.S. revenue on opening and opinion (outliers removed)

```
USRevenu = - 75.7 + 3.10 Opening + 11.1 Opinion

Predictor        Coef       Stdev    t-ratio        p
Constant       -75.72       14.44      -5.24    0.000
Opening       3.10379     0.07928      39.15    0.000
Opinion        11.112       2.341       4.75    0.000

s = 13.17       R-sq = 98.6%     R-sq(adj) = 98.5%
```

−2	0
−1	6
−1	420
−0	98765
−0	433210
0	0012334
0	56789
1	024
1	6
2	0



11.27. **(a)** The PEER distribution is left-skewed; the other two distributions are irregular (stemplots below). Student choices of summary statistics may vary; both five-number summaries and means/standard deviations are given below. **(b)** Correlation coefficients are given below the scatterplots. PEER and FtoS are negatively correlated, FtoS and CtoF are positively correlated, and the other correlation is very small.

Variable	$\bar{x}$	s	Min	Q_1	M	Q_3	Max
Peer review score	79.60	18.37	39	61	85	97	100
Faculty/student ratio	61.88	28.23	18	29	67	89	100
Citations/faculty ratio	63.84	25.23	17	40	66	86	100

Peer review score

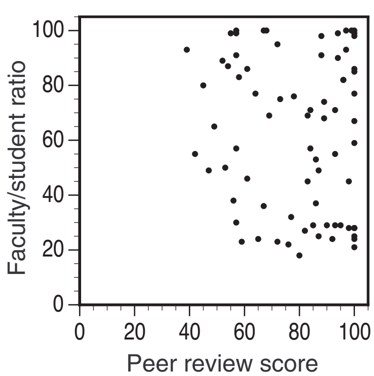
```
 3 | 9
 4 | 2
 4 | 579
 5 | 234
 5 | 567777789
 6 | 114
 6 | 57789
 7 | 223
 7 | 678
 8 | 023344
 8 | 566778899
 9 | 0233344
 9 | 56677889
10 | 00000000000000
```

Faculty/student ratio

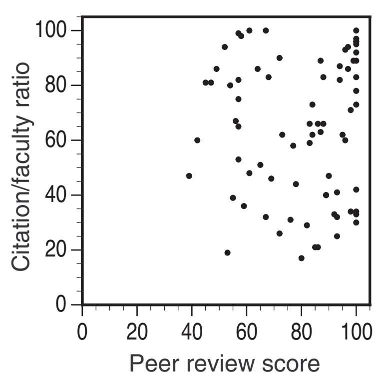
```
 1 | 8
 2 | 1233444
 2 | 55788889999
 3 | 02
 3 | 678
 4 |
 4 | 55699
 5 | 03
 5 | 55779
 6 |
 6 | 57899
 7 | 114
 7 | 5677
 8 | 0223
 8 | 56679
 9 | 01133
 9 | 5889999
10 | 000000
```

Citations/faculty ratio

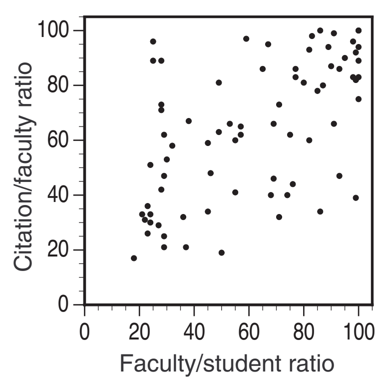
```
 1 | 79
 2 | 11
 2 | 569
 3 | 01223344
 3 | 69
 4 | 00124
 4 | 6778
 5 | 13
 5 | 89
 6 | 002223
 6 | 56667
 7 | 133
 7 | 58
 8 | 01122333
 8 | 6667999
 9 | 02344
 9 | 566789
10 | 000
```



Correlation: –0.1143



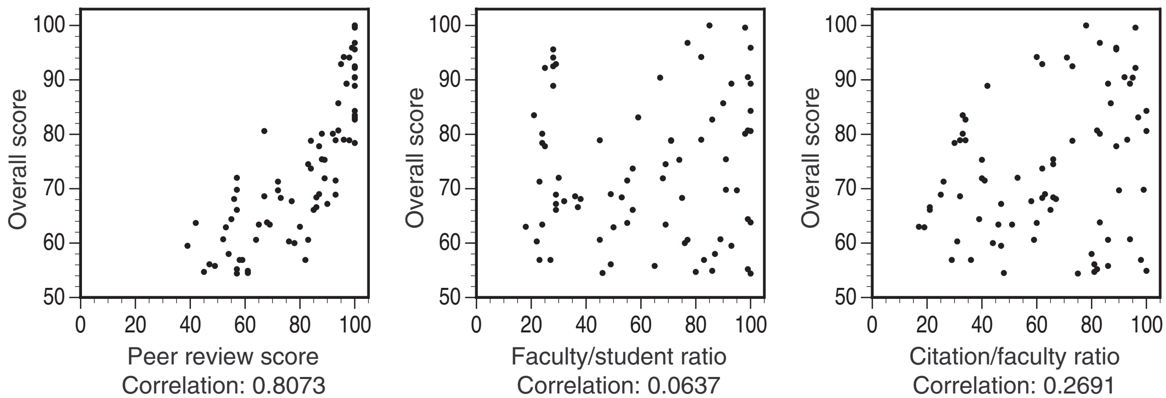
Correlation: 0.0045



Correlation: 0.5801

11.28. **(a)** All three scatterplots are on the following page. The plot versus peer review score is much more linear than the other two. **(b)** The correlations are given below the scatterplots. Not surprisingly, the PEER correlation is greatest.

Note: *The fact that the scatterplots do not all suggest linear associations does not mean that a multiple regression is inappropriate. Even if the data exactly fit a multiple regression model, the pairwise scatterplots will not necessarily appear to be linear.*



11.29. **(a)** The model is $\text{OVERALL}_i = \beta_0 + \beta_1\,\text{PEER}_i + \beta_2\,\text{FtoS}_i + \beta_3\,\text{CtoS}_i + \epsilon_i$, where ϵ_i are independent $N(0, \sigma)$ random variables. **(b)** The regression equation is:

$$\widehat{\text{OVERALL}} = 18.85 + 0.5746\,\text{PEER} + 0.0013\,\text{FtoS} + 0.1369\,\text{CtoF}$$

(c) For the confidence intervals, take $b_i \pm t^*\text{SE}_{b_i}$, with $t^* = 1.9939$ (for df $= 71$). These intervals have been added to the Minitab output below. The second interval contains 0, because that coefficient is not significantly different from 0. **(d)** The regression explains $R^2 \doteq 72.2\%$ of the variation in overall score. The estimate of σ is $s \doteq 7.043$.

Minitab output: Regression of overall score on all three variables

```
OVERALL = 18.8 + 0.575 PEER + 0.0013 FtoS + 0.137 CtoF

Predictor       Coef       Stdev    t-ratio        p   95% confidence interval
Constant      18.846       4.363       4.32    0.000
PEER         0.57462     0.04504      12.76    0.000     0.4848 to 0.6644
FtoS         0.00130     0.03597       0.04    0.971    -0.0704 to 0.0730
CtoF         0.13690     0.03999       3.42    0.001     0.0572 to 0.2166

s = 7.043        R-sq = 72.2%     R-sq(adj) = 71.0%
```

11.30. **(a)** Between GPA and IQ, $r \doteq 0.634$ (so straight-line regression explains $r^2 = 40.2\%$ of the variation in GPA). Between GPA and self-concept, $r \doteq 0.542$ (so regression explains $r^2 = 29.4\%$ of the variation in GPA). Since gender is categorical, the correlation between GPA and gender is not meaningful. **(b)** The model is $\text{GPA} = \beta_0 + \beta_1\,\text{IQ} + \beta_2\,\text{SC} + \epsilon_i$, where ϵ_i are independent $N(0, \sigma)$ random variables. **(c)** Regression gives the equation $\widehat{\text{GPA}} = -3.88 + 0.0772\,\text{IQ} + 0.0513\,\text{SC}$. Based on the reported value of R^2, the regression explains 47.1% of the variation in GPA. (So adding self-concept to the model only adds about 6.9% to the variation explained by the regression.) **(d)** We test H_0: $\beta_2 = 0$ versus H_a: $\beta_2 \neq 0$. The test statistic $t = 3.14$ (df $= 75$) has $P = 0.002$; we conclude that the coefficient of self-concept is not 0.

Minitab output: Regression of GPA on IQ and self-concept score

```
GPA = -3.88 + 0.0772 IQ + 0.0513 SelfCcpt

Predictor       Coef       Stdev    t-ratio        p
Constant      -3.882       1.472      -2.64    0.010
IQ           0.07720     0.01539       5.02    0.000
SelfCcpt     0.05125     0.01633       3.14    0.002

s = 1.547        R-sq = 47.1%     R-sq(adj) = 45.7%
```

11.31. (a) All distributions are skewed to varying degrees—GINI and CORRUPT to the right, the other three to the left. CORRUPT, DEMOCRACY, and LIFE have the most skewness. Student choices of summary statistics may vary; five-number summaries are a good choice because of the skewness, but some may also give means and standard deviations.

Variable	$\bar{x}$	s	Min	Q_1	M	Q_3	Max
LSI	6.2597	1.2773	2.5	5.5	6.1	7.35	8.2
GINI	37.9399	8.8397	24.70	32.65	35.95	42.750	60.10
CORRUPT	4.8861	2.4976	1.7	2.85	4.0	7.3	9.7
DEMOCRACY	4.2917	1.6799	0.5	3.0	5.0	5.5	6.0
LIFE	71.9450	9.0252	44.28	70.39	73.16	78.765	82.07

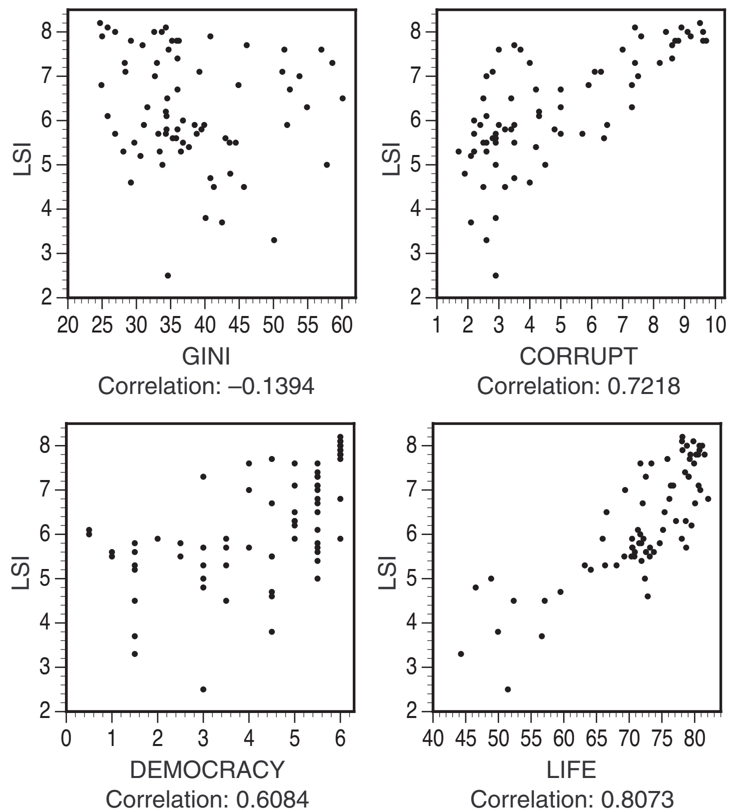
Notice especially how the skewness is apparent in the five-number summaries.
(b) Correlation coefficients are given below the scatterplots. GINI is negatively (and weakly) correlated to the other four variables, while all other correlations are positive and more substantial (0.533 or more).

LSI

```
2 | 5
3 | 3
3 | 78
4 |
4 | 55678
5 | 0023334
5 | 555566677778889999
6 | 011233
6 | 557788
7 | 001113334
7 | 66677888899
8 | 000112
```

GINI

```
2 | 44
2 | 55566888999
3 | 0011223333344444444
3 | 55556666666788999
4 | 0001233344
4 | 56
5 | 0112234
5 | 778
6 | 0
```



CORRUPT

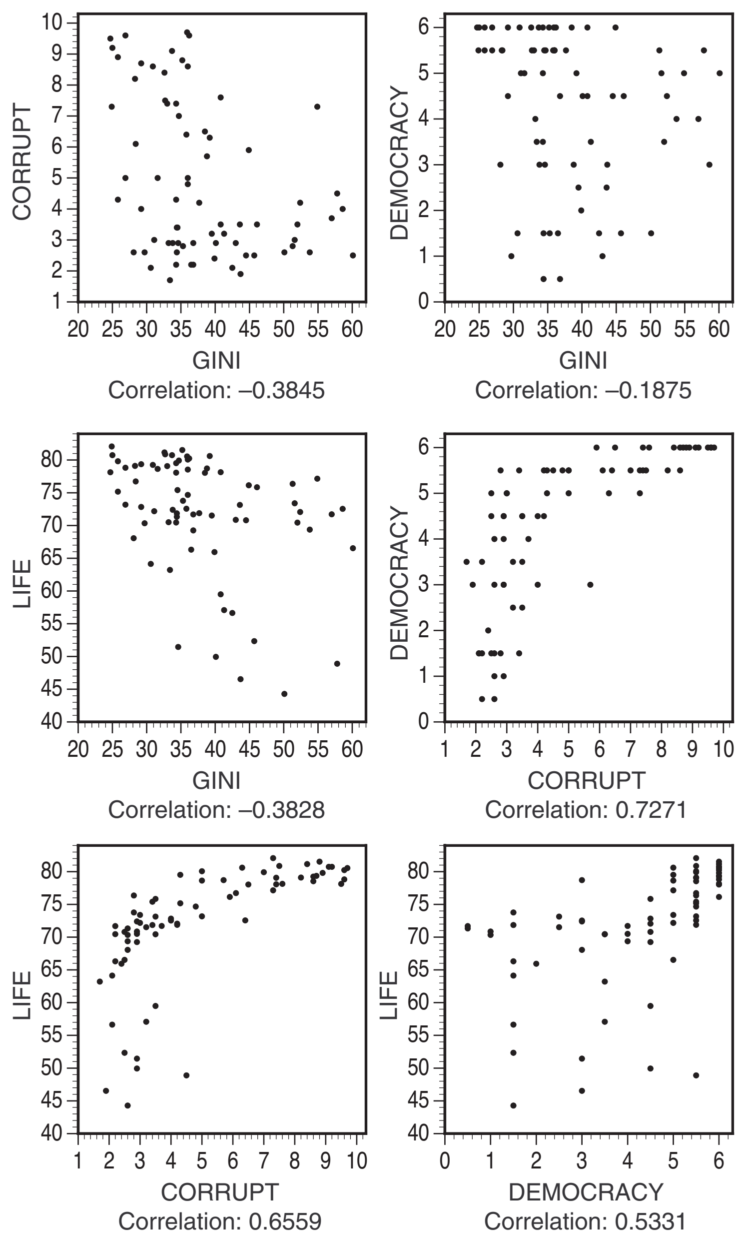
1	79
2	112224
2	5556666688999999
3	002244
3	55557
4	002233
4	58
5	000
5	79
6	134
6	5
7	03344
7	56
8	24
8	66789
9	12
9	5667

DEMOCRACY

0	55
1	00
1	5555555
2	0
2	55
3	000000
3	5555
4	000
4	5555555
5	0000000
5	5555555555555555
6	00000000000000

LIFE

4	4
4	689
5	12
5	679
6	34
6	566899
7	000000111111222222233334
7	55566678888888889999999
8	0000000112



11.32. The four regression equations are:

(**a**) $\widehat{\text{LSI}} = \quad 7.02 - 0.0201\,\text{GINI}$

(**b**) $\widehat{\text{LSI}} = -3.83 + 0.0287\,\text{GINI} + 0.1250\,\text{LIFE}$

(**c**) $\widehat{\text{LSI}} = -3.25 + 0.0280\,\text{GINI} + 0.1063\,\text{LIFE} + 0.1857\,\text{DEMOCRACY}$

(**d**) $\widehat{\text{LSI}} = -2.72 + 0.0368\,\text{GINI} + 0.0905\,\text{LIFE} + 0.0392\,\text{DEMOCRACY} + 0.1855\,\text{CORRUPT}$

Minitab output (on the next page) gives the values of R^2 for each regression and highlights non-significant P-values. We note that GINI does not contribute significantly to the first model but is significant in every other model, and DEMOCRACY is not significant in the last model, even though it was significant in the second-to-last model. (Roughly speaking, this means that whatever information DEMOCRACY contributed to the model, CORRUPT contains that same information but contributes it more efficiently than does DEMOCRACY. Recall from the previous solution that DEMOCRACY and CORRUPT had correlation 0.7271.)

Shown on the next page are stemplots of the residuals for all four regressions; the first distribution is clearly skewed, but the other three show no severe deviations from Normality. A full analysis of the residuals for each regression would require a total of 10 scatterplots; shown are six of these plots which suggest possible problems with the assumptions. The first five show signs of non-constant standard deviations, and the last shows a hint of curvature.

Minitab output: Regression of LSI on GINI (Model 1)

```
LSI = 7.02 - 0.0201 GINI

Predictor       Coef       Stdev    t-ratio       p
Constant      7.0238      0.6660      10.55   0.000
GINI        -0.02014     0.01710      -1.18   0.243

s = 1.274       R-sq = 1.9%     R-sq(adj) = 0.5%
```

Regression of LSI on GINI and LIFE (Model 2)

```
LSI = - 3.83 + 0.0287 GINI + 0.125 LIFE

Predictor       Coef       Stdev    t-ratio       p
Constant     -3.8257      0.9746      -3.93   0.000
GINI         0.02873     0.01056       2.72   0.008
LIFE         0.12503     0.01034      12.09   0.000

s = 0.7266      R-sq = 68.6%    R-sq(adj) = 67.6%
```

Regression of LSI on GINI, LIFE, and DEMOCRACY (Model 3)

```
LSI = - 3.25 + 0.0280 GINI + 0.106 LIFE + 0.186 DEMOCRACY

Predictor       Coef       Stdev    t-ratio       p
Constant     -3.2524      0.9293      -3.50   0.001
GINI        0.028049    0.009891       2.84   0.006
LIFE         0.10634     0.01125       9.46   0.000
DEMOCRACY    0.18575     0.05682       3.27   0.002

s = 0.6804      R-sq = 72.8%    R-sq(adj) = 71.6%
```

Regression of LSI on all four variables (Model 4)

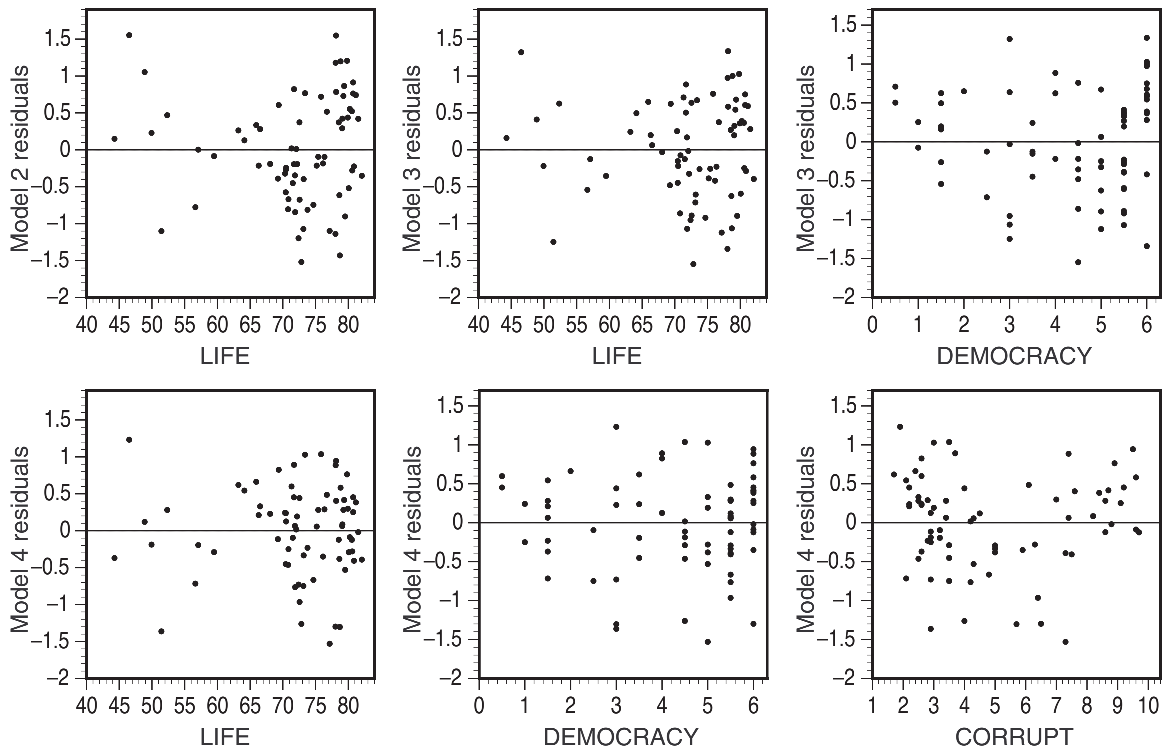
```
LSI = - 2.72 + 0.0368 GINI + 0.0905 LIFE + 0.0392 DEMOCRACY + 0.186 CORRUPT

Predictor       Coef       Stdev    t-ratio       p
Constant     -2.7201      0.8661      -3.14   0.003
GINI        0.036782    0.009393       3.92   0.000
LIFE         0.09048     0.01120       8.08   0.000
DEMOCRACY    0.03925     0.06566       0.60   0.552
CORRUPT      0.18554     0.05042       3.68   0.000

s = 0.6252      R-sq = 77.4%    R-sq(adj) = 76.0%
```

Model 1 residuals	
−3	8
−3	
−2	7
−2	44
−1	8665
−1	33210
−0	998877776666555
−0	44443322100
0	1234
0	66667889
1	0112233444
1	55566666677

Model 2 residuals	
−1	54
−1	
−1	11100
−0	9888
−0	77666
−0	554
−0	33333222222
−0	1111000
0	00011
0	2222333
0	4444555
0	6777777
0	889
1	011
1	2
1	55

Model 3 residuals	
−1	5
−1	32
−1	100
−0	99888
−0	766
−0	55444
−0	33332222222
−0	111000
0	01111
0	222233333
0	445555
0	6666666777
0	89
1	00
1	33

Model 4 residuals	
−1	5
−1	3322
−1	
−0	9
−0	77776
−0	5444
−0	3333322222
−0	11111000
0	00000111
0	222222222233
0	444444555
0	667
0	8889
1	00
1	2



11.33. **(a)** The coefficients, standard errors, t statistics, and P-values are given in the Minitab output shown with the solution to the previous exercise. **(b)** Student observations will vary. For example, the t statistic for the GINI coefficient grows from $t = -1.18$ ($P = 0.243$) to $t = 3.92$ ($P < 0.0005$). The DEMOCRACY t is 3.27 in the third model ($P < 0.0005$) but drops to 0.60 ($P = 0.552$) in the fourth model. **(c)** A good choice is to use GINI, LIFE, and CORRUPT (Minitab output on the following page). All three coefficients are significant, and $R^2 = 77.3\%$ is nearly the same as the fourth model from previous exercise. However, a scatterplot of the residuals versus CORRUPT (not shown) still looks quite a bit like the final scatterplot shown in the previous solution, suggesting a slightly curved relationship, which would violate the assumptions of our model.

Minitab output: Regression of LSI on GINI, LIFE, and CORRUPT

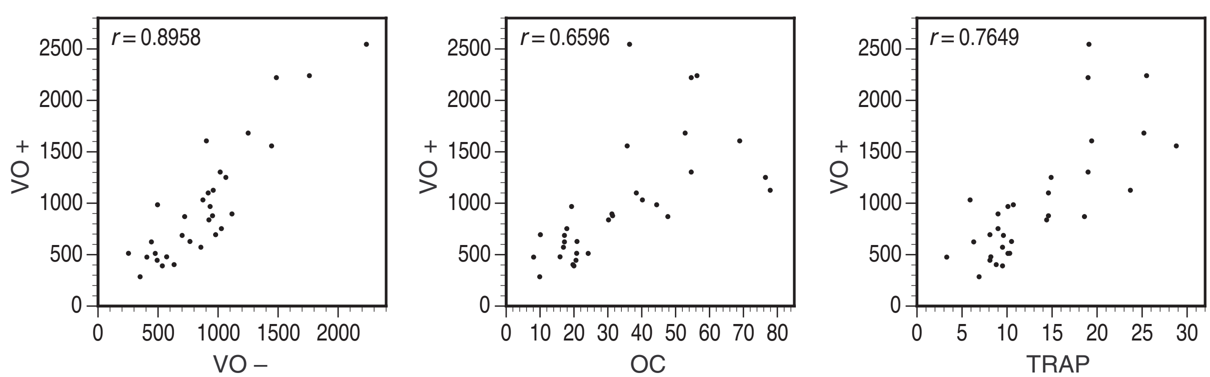
```
LSI = - 2.74 + 0.0377 GINI + 0.0914 LIFE + 0.204 CORRUPT

Predictor       Coef       Stdev    t-ratio        p
Constant     -2.7442      0.8610      -3.19    0.002
GINI        0.037734    0.009213       4.10    0.000
LIFE         0.09141     0.01104       8.28    0.000
CORRUPT      0.20382     0.03991       5.11    0.000

s = 0.6222      R-sq = 77.3%     R-sq(adj) = 76.3%
```

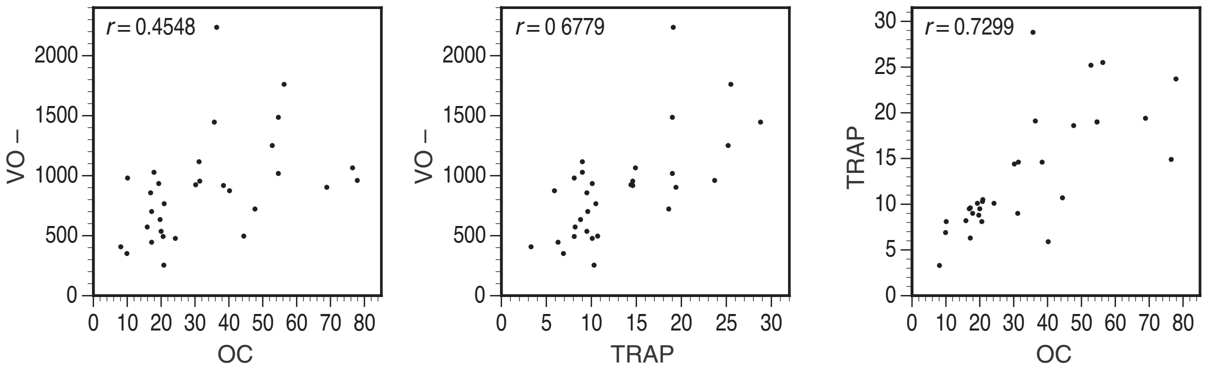
11.34. (a) Stemplots (below) show that all four variables are right-skewed to some degree.

Variable	$\bar{x}$	s	Min	Q_1	M	Q_3	Max
VO+	985.806	579.858	285.0	513.0	870.0	1251.0	2545.0
VO–	889.194	427.616	254.0	536.0	903.0	1028.0	2236.0
OC	33.416	19.610	8.1	17.9	30.2	47.7	77.9
TRAP	13.248	6.528	3.3	8.8	10.3	19.0	28.8

(b) Correlations and scatterplots (below) show that all six pairs of variables are positively associated. The strongest association is between VO+ and VO– and the weakest is between OC and VO–.

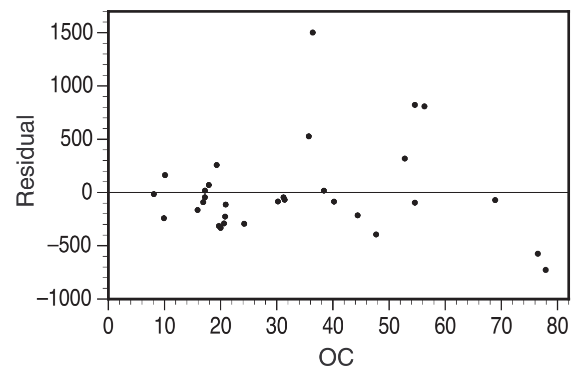
```
VO+            VO–             OC             TRAP
0 | 23         0 | 23          0 | 89         0 | 3
0 | 4444555    0 | 4444455     1 | 0          0 | 5
0 | 66667      0 | 6777        1 | 5677799    0 | 66
0 | 888899     0 | 889999999   2 | 00004      0 | 888899999
1 | 011        1 | 0001        2 |            1 | 00000
1 | 23         1 | 2           3 | 011        1 |
1 | 5          1 | 44          3 | 568        1 | 4444
1 | 66         1 | 7           4 | 04         1 |
1 |            1 |             4 | 7          1 | 89999
2 |            2 |             5 | 244        2 |
2 | 22         2 | 2           5 | 6          2 | 3
2 | 5                          6 |            2 | 55
                               6 | 8          2 |
                               7 |            2 | 8
                               7 | 67
```





11.35. (a) See the previous solution for the scatterplot, which suggests greater variation in VO+ for large OC. The regression equation is

$$\widehat{\text{VO+}} = 334.0 + 19.505\ \text{OC}$$



with $s \doteq 443.3$ and $R^2 \doteq 0.435$; the test statistic for the slope is $t = 4.73$ ($P < 0.0005$), so we conclude the slope is not zero. The plot of residuals against OC suggests a slight downward curve on the right end, as well as increasing scatter as OC increases. The residuals are also somewhat right-skewed. A stemplot and Normal quantile plot of the residuals are not shown here but could be included as part of the analysis. **(b)** The regression equation is

$$\widehat{\text{VO+}} = 57.7 + 6.415\ \text{OC} + 53.87\ \text{TRAP}$$

with $s \doteq 376.3$ and $R^2 \doteq 0.607$. The coefficient of OC is not significantly different from 0 ($t = 1.25$, $P = 0.221$), but the coefficient of TRAP is ($t = 3.50$, $P = 0.002$). This is consistent with the correlations found in the solution to Exercise 11.34: TRAP is more highly correlated with VO+, and is also highly correlated with OC, so it is reasonable that, if TRAP is present in the model, little additional information is gained from OC.

Minitab output: Regression of VO+ on OC (Model 1)

```
The regression equation is VOplus = 334 + 19.5 OC

Predictor        Coef        Stdev    t-ratio        p
Constant        334.0        159.2       2.10    0.045
OC             19.505        4.127       4.73    0.000

s = 443.3       R-sq = 43.5%     R-sq(adj) = 41.6%
```

Regression of VO+ on OC and TRAP (Model 2)

```
The regression equation is VOplus = 58 + 6.41 OC + 53.9 TRAP

Predictor        Coef        Stdev    t-ratio        p
Constant         57.7        156.5       0.37    0.715
OC              6.415        5.125       1.25    0.221
TRAP            53.87        15.39       3.50    0.002

s = 376.3       R-sq = 60.7%     R-sq(adj) = 57.9%
```

11.36. **(a)** The model is

$$\text{VO+} = \beta_0 + \beta_1\ \text{OC} + \beta_2\ \text{TRAP} + \beta_2\ \text{VOminus} + \epsilon_i$$

where ϵ_i are independent $N(0, \sigma)$ random variables. **(b)–(d)** The table below summarizes the results for all the regressions called for in this exercise. The estimated coefficients and P-values can change rather drastically from one model to the next. Generally, R^2 increases (sometimes only slightly) as we add more explanatory variables to the model. **(e)** The results of the regression in part (b) suggest that we remove TRAP from the model. This regression equation and associated results are also in the table below. Because R^2 drops only slightly for this simpler model, this is probably the best of all models we have considered to this point.

Model						R^2	s
1	$\widehat{\text{VO+}} =$	334.0	$+$ 19.505 OC			0.435	443.3
			SE $= 4.127$				
			$t = 4.73$				
			$P < 0.0005$				
2	$\widehat{\text{VO+}} =$	57.7	$+$ 6.415 OC	$+$ 53.87 TRAP		0.607	376.3
			SE $= 5.125$	SE $= 15.39$			
			$t = 1.25$	$t = 3.50$			
			$P = 0.221$	$P = 0.002$			
3	$\widehat{\text{VO+}} =$	-243.5	$+$ 8.235 OC	$+$ 6.61 TRAP	$+$ 0.975 VOminus	0.884	207.8
			SE $= 2.840$	SE $= 10.33$	SE $= 0.1211$		
			$t = 2.90$	$t = 0.64$	$t = 8.05$		
			$P = 0.007$	$P = 0.528$	$P < 0.0005$		
4	$\widehat{\text{VO+}} =$	-234.1	$+$ 9.404 OC		$+$ 1.019 VOminus	0.883	205.6
			SE $= 2.150$		SE $= 0.0986$		
			$t = 4.37$		$t = 10.33$		
			$P < 0.0005$		$P < 0.0005$		

Minitab output: Regression of VO+ on OC, TRAP and VO– (Model 3)

```
The regression equation is VOplus = -243 + 8.23 OC + 6.6 TRAP + 0.975 VOminus

Predictor       Coef        Stdev    t-ratio        p
Constant     -243.49        94.22      -2.58    0.015
OC             8.235        2.840       2.90    0.007
TRAP            6.61        10.33       0.64    0.528
VOminus       0.9746       0.1211       8.05    0.000

s = 207.8       R-sq = 88.4%     R-sq(adj) = 87.2%
```

Regression of VO+ on OC and VO– (Model 4)

```
The regression equation is VOplus = -234 + 9.40 OC + 1.02 VOminus

Predictor       Coef        Stdev    t-ratio        p
Constant     -234.14        92.09      -2.54    0.017
OC             9.404        2.150       4.37    0.000
VOminus      1.01857      0.09858      10.33    0.000

s = 205.6       R-sq = 88.3%     R-sq(adj) = 87.4%
```

11.37. Stemplots (below) show that all four variables are noticeably less skewed.

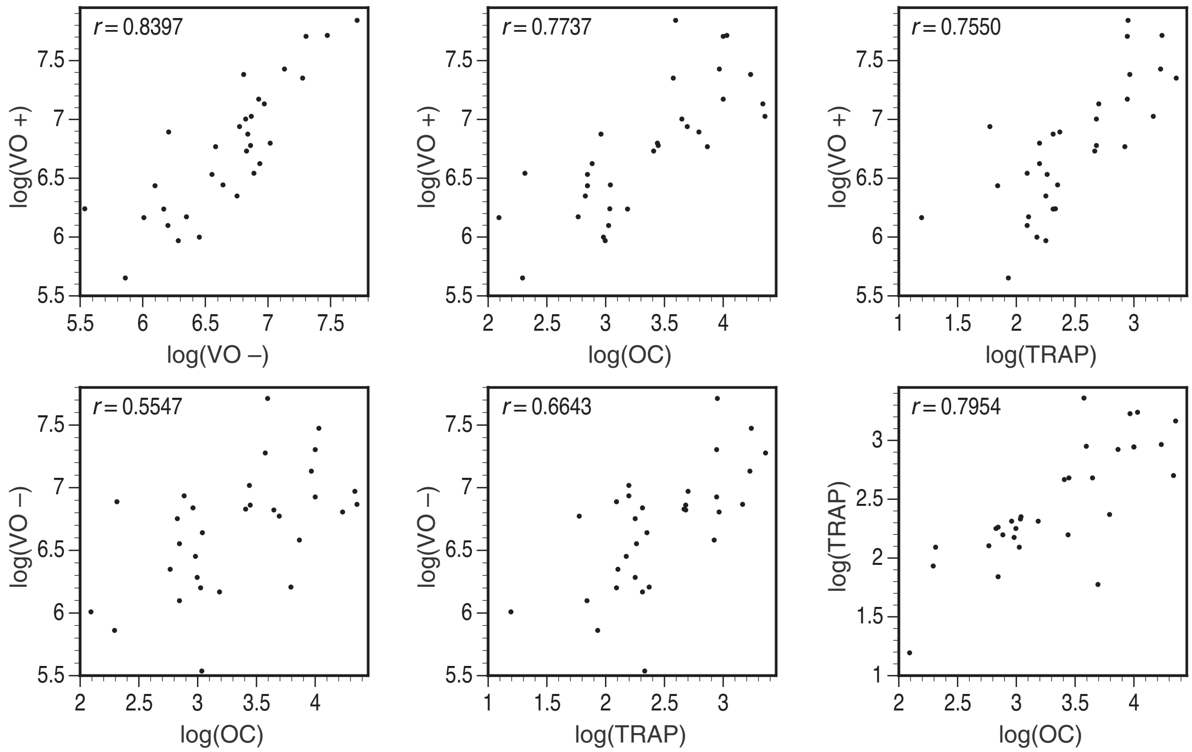
Variable	$\bar{x}$	s	Min	Q_1	M	Q_3	Max
LVO+	6.7418	0.5555	5.652	6.240	6.768	7.132	7.842
LVO–	6.6816	0.4832	5.537	6.284	6.806	6.935	7.712
LOC	3.3380	0.6085	2.092	2.885	3.408	3.865	4.355
LTRAP	2.4674	0.4978	1.194	2.175	2.332	2.944	3.360

Correlations and scatterplots (on the following page) show that all six pairs of variables are positively associated. The strongest association is between LVO+ and LVO– and the weakest is between LOC and LVO–. The regression equations for these transformed variables are given in the table below, along with significance test results. Residual analysis for these regressions is not shown.

The final conclusion is the same as for the untransformed data: When we use all three explanatory variables to predict LVO+, the coefficient of LTRAP is not significantly different from 0 and we then find that the model that uses LOC and LVO– to predict LVO+ is nearly as good (in terms of R^2), making it the best of the bunch.

LVO+		LVO–		LOC		LTRAP	
5	6	5	5	2	0	1	1
5	99	5		2	23	1	
6	011	5	8	2		1	
6	223	6	001	2	7	1	7
6	4455	6	2223	2	8888999	1	89
6	67777	6	455	3	0001	2	001111
6	889	6	677	3		2	22233333
7	0011	6	8888888999	3	44455	2	
7	33	7	01	3	667	2	6667
7	4	7	23	3	89	2	99999
7	77	7	4	4	000	3	1
7	8	7	7	4	233	3	223

					R^2	s
$\widehat{\text{LVO+}} = 4.3841$	$+\,0.7063$ LOC				0.599	0.3580
	SE = 0.1074 $t = 6.58$ $P < 0.0005$					
$\widehat{\text{LVO+}} = 4.2590$	$+\,0.4304$ LOC	$+\,0.4240$ LTRAP			0.652	0.3394
	SE = 0.1680 $t = 2.56$ $P = 0.016$	SE = 0.2054 $t = 2.06$ $P = 0.048$				
$\widehat{\text{LVO+}} = 0.8716$	$+\,0.3922$ LOC	$+\,0.0275$ LTRAP	$+\,0.6725$ LVOminus		0.842	0.2326
	SE = 0.1154 $t = 3.40$ $P = 0.002$	SE = 0.1570 $t = 0.18$ $P = 0.842$	SE = 0.1178 $t = 5.71$ $P < 0.0005$			
$\widehat{\text{LVO+}} = 0.8321$	$+\,0.4061$ LOC		$+\,0.6816$ LVOminus		0.842	0.2286
	SE = 0.0824 $t = 4.93$ $P < 0.0005$		SE = 0.1038 $t = 6.57$ $P < 0.0005$			



11.38. Refer to the solution to Exercise 11.34 for the scatterplots. Note that, in this case, it really makes the most sense to use TRAP (rather than OC) to predict VO– (because it is the appropriate biomarker), but many students might miss that detail. Both single-explanatory variable models are given in the first table on the following page. Residual analysis plots are not included. Our conclusion here is similar to the conclusion in Exercises 11.36 and 11.37: The best model is to use OC and VO+ to predict VO–.

11.39. Refer to the solution to Exercise 11.37 for the scatterplots. As in the previous exercise, the more logical single-variable model would be to use LTRAP to predict LVO–, but many students might miss that detail. Both single-explanatory variable models are given in the second table on the following page. Residual analysis plots are not included. This time, we might conclude that the best model is to predict LVO– from LVO+ alone; neither biomarker variable makes an indispensable contribution to the prediction.

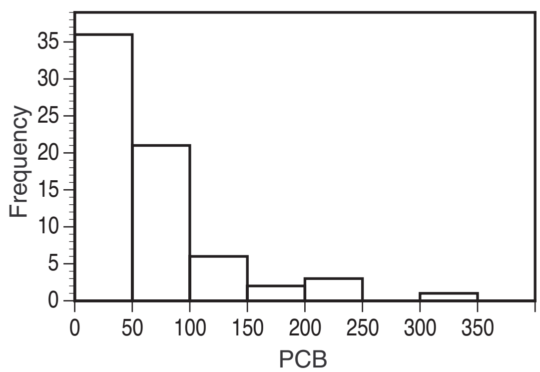
For Exercise 11.38					R^2	s
$\widehat{\text{VO}-} =$	557.8	$+ 9.917$ OC SE $= 3.606$ $t = 2.75$ $P = 0.010$			0.207	387.4
$\widehat{\text{VO}-} =$	300.9		$+ 44.41$ TRAP SE $= 8.942$ $t = 4.97$ $P < 0.0005$		0.460	319.7
$\widehat{\text{VO}-} =$	309.1	$- 1.868$ OC SE $= 4.418$ $t = -0.42$ $P = 0.676$	$+ 48.50$ TRAP SE $= 13.27$ $t = 3.66$ $P = 0.001$		0.463	324.4
$\widehat{\text{VO}-} =$	267.3	$- 6.513$ OC SE $= 2.507$ $t = -2.60$ $P = 0.015$	$+ 9.485$ TRAP SE $= 8.788$ $t = 1.08$ $P = 0.029$	$+ 0.724$ VOplus SE $= 0.090$ $t = 8.05$ $P < 0.0005$	0.842	179.2
$\widehat{\text{VO}-} =$	298.0	$- 5.254$ OC SE $= 2.226$ $t = -2.36$ $P = 0.025$		$+ 0.778$ VOplus SE $= 0.0753$ $t = 10.33$ $P < 0.0005$	0.835	179.7

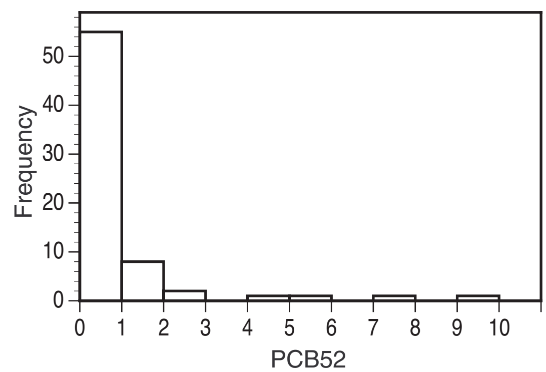
For Exercise 11.39					R^2	s
$\widehat{\text{LVO}-} =$	5.2110	$+ 0.4406$ LOC SE $= 0.1227$ $t = 3.59$ $P = 0.001$			0.308	0.4089
$\widehat{\text{LVO}-} =$	5.0905		$+ 0.6449$ LTRAP SE $= 0.1347$ $t = 4.79$ $P < 0.0005$		0.441	0.3674
$\widehat{\text{LVO}-} =$	5.0370	$+ 0.0569$ LOC SE $= 0.1848$ $t = 0.31$ $P = 0.761$	$+ 0.5896$ LTRAP SE $= 0.2259$ $t = 2.61$ $P = 0.014$		0.443	0.3732
$\widehat{\text{LVO}-} =$	1.5729	$- 0.2932$ LOC SE $= 0.1407$ $t = -2.08$ $P = 0.047$	$+ 0.2447$ LTRAP SE $= 0.1662$ $t = 1.47$ $P = 0.152$	$+ 0.8134$ LVOplus SE $= 0.1425$ $t = 5.71$ $P < 0.0005$	0.748	0.2558
$\widehat{\text{LVO}-} =$	1.3109	$- 0.1878$ LOC SE $= 0.1237$ $t = -1.52$ $P = 0.140$		$+ 0.8896$ LVOplus SE $= 0.1355$ $t = 6.57$ $P < 0.0005$	0.728	0.2611
$\widehat{\text{LVO}-} =$	1.7570			$+ 0.7304$ LVOplus SE $= 0.0877$ $t = 8.33$ $P < 0.0005$	0.705	0.2669

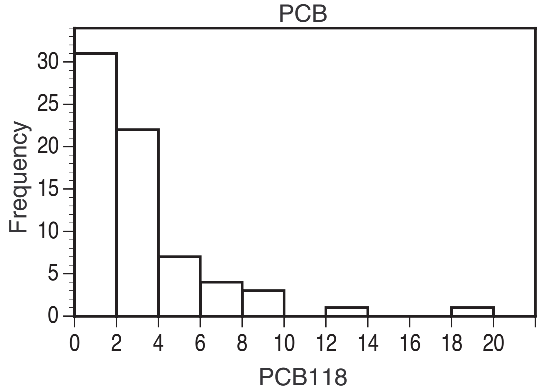
11.40. (a) Histograms are below; all distributions are sharply right-skewed.

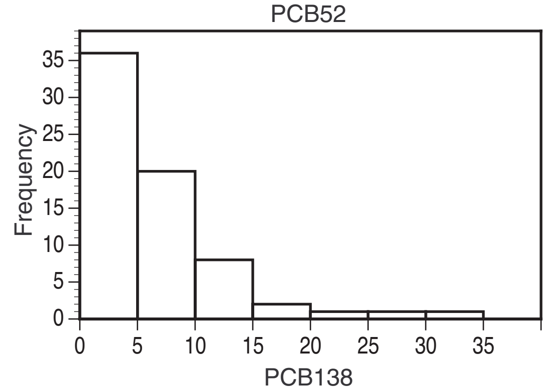
Variable	$\bar{x}$	s	Min	Q_1	M	Q_3	Max
PCB	68.4674	59.3906	6.0996	29.8305	47.9596	91.7140	318.746
PCB52	0.9580	1.5983	0.0200	0.2180	0.4770	0.8925	9.060
PCB118	3.2563	3.0191	0.2360	1.4800	2.4200	3.8950	18.900
PCB138	6.8268	5.8627	0.6400	2.9700	4.9200	8.7150	32.300
PCB180	4.1584	4.9864	0.3950	1.1950	2.6900	4.5900	31.500

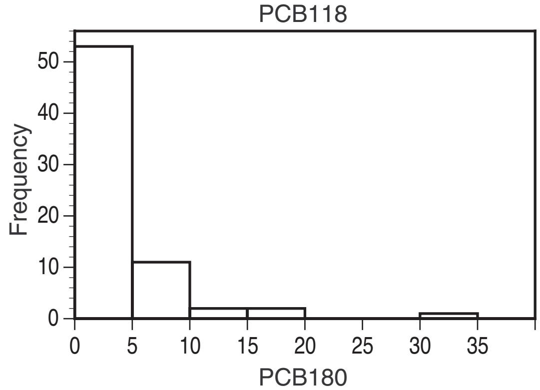
(b) Scatterplots and correlations are on the following page. All pairs of variables are positively associated, although some only weakly. In general, even when the association is strong, the plots show more variation for large values of the two variables. If we test H_0: $\rho = 0$ versus H_a: $\rho \neq 0$ for these correlations, we find that $P < 0.0005$ for eight of them. The PCB52/PCB138 correlation is less significant ($r \doteq 0.3009$, $t = 2.58$, $P = 0.0120$), and the PCB52/PCB180 correlation is not significantly different from 0 ($r \doteq 0.0869$, $t = 0.71$, $P = 0.4775$).

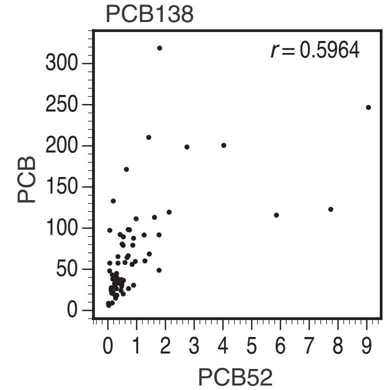


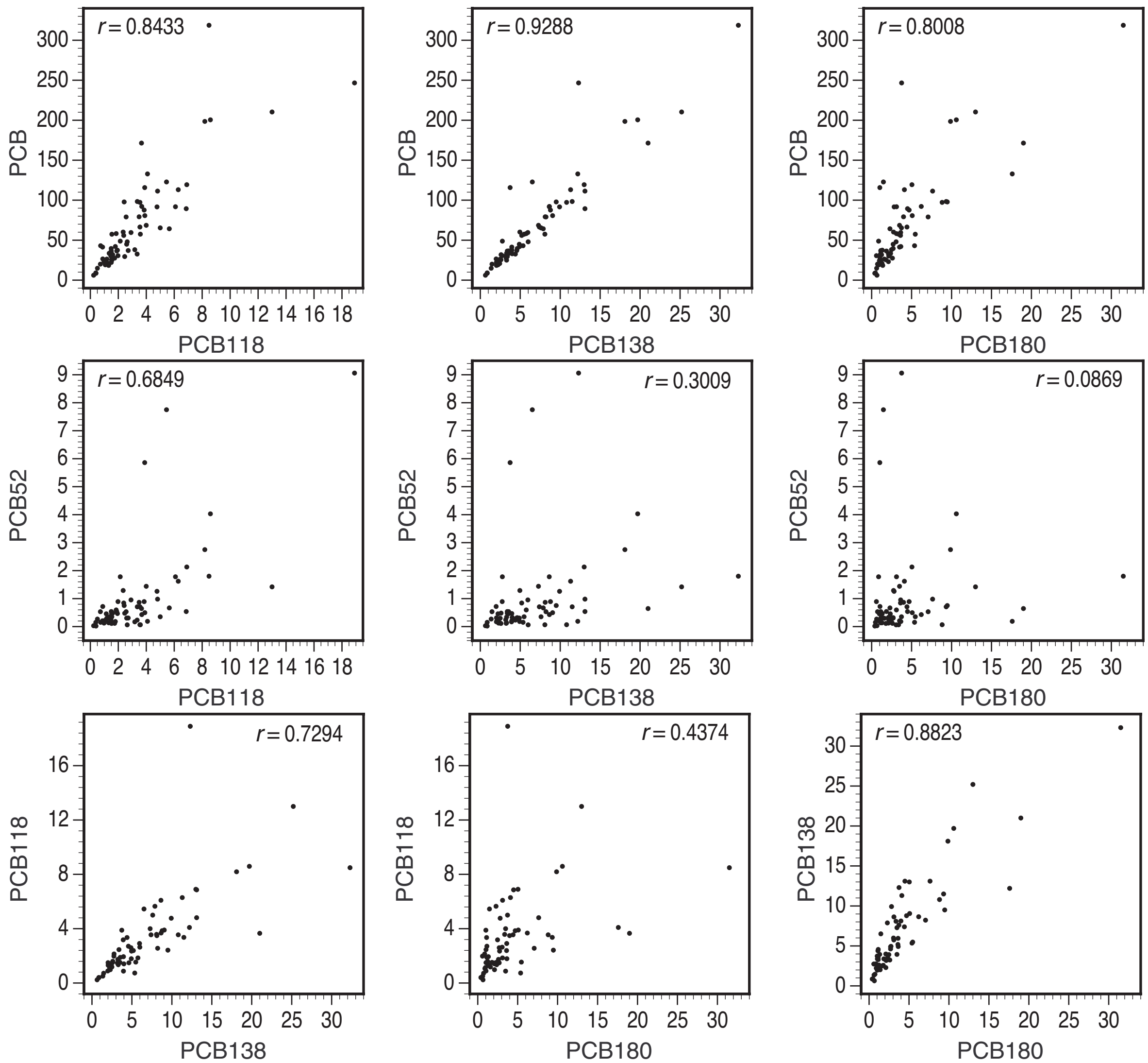












11.41. **(a)** The model is:

$$\text{PCB}_i = \beta_0 + \beta_1\,\text{PCB}_{52} + \beta_2\,\text{PCB}_{118} + \beta_3\,\text{PCB}_{138} + \beta_4\,\text{PCB}_{180} + \epsilon_i$$

where $i = 1, 2, \ldots, 69$; ϵ_i are independent $N(0, \sigma)$ random variables. **(b)** The regression equation is:

$$\widehat{\text{PCB}} = 0.937 + 11.8727\,\text{PCB}_{52} + 3.7611\,\text{PCB}_{118} + 3.8842\,\text{PCB}_{138} + 4.1823\,\text{PCB}_{180}$$

−2	2
−1	
−1	31
−0	8776655
−0	4443333222211111000000
0	00000000000111112222223333444
0	677778
1	12
1	
2	2

with $s = 6.382$ and $R^2 = 0.989$. All coefficients are significantly different from 0, although the constant 0.937 is not ($t = 0.76$, $P = 0.449$). That makes some sense—if none of these four congeners are present, it might be somewhat reasonable to predict that the total amount of PCB is 0. **(c)** The residuals appear to be roughly Normal, but with two outliers. There are no clear patterns when plotted against the explanatory variables (these plots are not shown).

Minitab output: Regression of PCB on PCB52, PCB118, PCB138, and PCB180

```
PCB = 0.94 + 11.9 PCB52 + 3.76 PCB118 + 3.88 PCB138 + 4.18 PCB180

Predictor       Coef       Stdev    t-ratio        p
Constant       0.937       1.229       0.76    0.449
PCB52        11.8727      0.7290      16.29    0.000
PCB118        3.7611      0.6424       5.85    0.000
PCB138        3.8842      0.4978       7.80    0.000
PCB180        4.1823      0.4318       9.69    0.000

s = 6.382       R-sq = 98.9%     R-sq(adj) = 98.8%
```

11.42. **(a)** The outliers are specimen #50 (residual −22.0864) and #65 (22.5487). Because residuals are observed values minus predicted values, the negative residual (#50) is an overestimate. (The estimated PCB for this specimen is $\widehat{\text{PCB}} \doteq 144.882$, and the actual level was 122.796.) **(b)** The regression equation is:

$$\widehat{\text{PCB}} = 1.6277 + 14.4420\,\text{PCB}_{52} + 2.5996\,\text{PCB}_{118} + 4.0541\,\text{PCB}_{138} + 4.1086\,\text{PCB}_{180}$$

with $s = 4.555$ and $R^2 = 0.994$. As before, all coefficients are significantly different from 0, although the constant is barely not different ($t = 1.84$, $P = 0.071$). The residuals again appear to be roughly Normal, but two new specimens (#44 and #58) show up as outliers to replace the two we removed. There are no clear patterns when plotted against the explanatory variables (these plots are not shown).

−1	2
−1	
−0	98
−0	76
−0	5544
−0	33332222
−0	1111111100000000000
0	0000000000111111
0	22233
0	445
0	677
0	889
1	
1	
1	4

Minitab output: Regression of PCB on the four congeners (outliers removed)

```
PCB = 1.63 + 14.4 PCB52 + 2.60 PCB118 + 4.05 PCB138 + 4.11 PCB180

Predictor       Coef       Stdev    t-ratio        p
Constant      1.6277      0.8858       1.84    0.071
PCB52        14.4420      0.6960      20.75    0.000
PCB118        2.5996      0.5164       5.03    0.000
PCB138        4.0541      0.3752      10.80    0.000
PCB180        4.1086      0.3175      12.94    0.000

s = 4.555       R-sq = 99.4%     R-sq(adj) = 99.4%
```

11.43. **(a)** The regression equation is:

$$\widehat{\text{PCB}} = -1.018 + 12.644\,\text{PCB}_{52} + 0.3131\,\text{PCB}_{118} + 8.2546\,\text{PCB}_{138}$$

with $s = 9.945$ and $R^2 = 0.973$. Residual analysis (not shown) suggests a few areas of concern: The distribution of residuals has heavier tails than a Normal distribution, and the scatter (that is, prediction error) is greater for larger values of the predicted PCB. **(b)** The estimated coefficient of PCB118 is $b_2 \doteq 0.3131$; its P-value is 0.708. (Details in Minitab output on the following page.) **(c)** In Exercise 11.41, $b_2 \doteq 3.7611$ and $P < 0.0005$. **(d)** This illustrates how complicated multiple regression can be: When we add PCB180 to the model, it complements PCB118, making it useful for prediction.

Minitab output: Regression of PCB on PCB52, PCB118, and PCB138

```
PCB = -1.02 + 12.6 PCB52 + 0.313 PCB118 + 8.25 PCB138

Predictor       Coef       Stdev    t-ratio       p
Constant      -1.018       1.890      -0.54   0.592
PCB52         12.644       1.129      11.20   0.000
PCB118        0.3131      0.8333       0.38   0.708
PCB138        8.2546      0.3279      25.18   0.000

s = 9.945       R-sq = 97.3%     R-sq(adj) = 97.2%
```

11.44. (a) Because TEQ is defined as the sum TEQPCB + TEQDIOXIN + TEQFURAN, we have $\beta_0 = 0$ and $\beta_1 = \beta_2 = \beta_3 = 1$. **(b)** The error terms are all zero, so they have no scatter; therefore, $\sigma = 0$. **(c)** Results will vary slightly with software, but except for rounding error, the regression confirms the values in parts (a) and (b).

Minitab output: Regression of TEQ on TEQPCB, TEQDIOXIN, and TEQFURAN

```
TEQ =0.000000 + 1.00 TEQPCB + 1.00 TEQDIOXIN + 1.00 TEQFURAN

Predictor       Coef       Stdev    t-ratio       p
Constant  0.00000032  0.00000192       0.16   0.870
TEQPCB       1.00000     0.00000 1211707.25   0.000
TEQDIOXIN    1.00000     0.00000  566800.75   0.000
TEQFURAN     1.00000     0.00001  176270.48   0.000

s = 0.000007964   R-sq = 100.0%    R-sq(adj) = 100.0%
```

11.45. The model is:

$$\text{TEQ}_i = \beta_0 + \beta_1\,\text{PCB}_{52} + \beta_2\,\text{PCB}_{118} + \beta_3\,\text{PCB}_{138} + \beta_4\,\text{PCB}_{180} + \epsilon_i$$

where $i = 1, 2, \ldots, 69$; ϵ_i are independent $N(0, \sigma)$ random variables. The regression equation is:

$$\widehat{\text{TEQ}} = 1.0600 - 0.0973\,\text{PCB}_{52} + 0.3062\,\text{PCB}_{118} + 0.1058\,\text{PCB}_{138} - 0.0039\,\text{PCB}_{180}$$

−1	66
−1	4200
−0	987666666666555555
−0	44444333221111100
0	0000222224
0	566667788
1	23334
1	9
2	3
2	57

with $s = 0.9576$ and $R^2 = 0.677$. Only the constant and the PCB118 coefficient are significantly different from 0; see Minitab output below. Residuals (stemplot on the right) are slightly right-skewed and show no clear patterns when plotted with the explanatory variables (not shown).

Minitab output: Regression of TEQ on the four PCB congeners

```
TEQ = 1.06 - 0.097 PCB52 + 0.306 PCB118 + 0.106 PCB138 - 0.0039 PCB180

Predictor       Coef       Stdev    t-ratio       p
Constant      1.0600      0.1845       5.75   0.000
PCB52        -0.0973      0.1094      -0.89   0.377
PCB118       0.30618     0.09639       3.18   0.002
PCB138       0.10579     0.07470       1.42   0.162
PCB180      -0.00391     0.06478      -0.06   0.952

s = 0.9576       R-sq = 67.7%     R-sq(adj) = 65.7%
```

11.46. **(a)** Results will vary with software. **(b)** Different software may produce different results, but (presumably) all software will ignore those 16 specimens, which is probably not a good approach. **(c)** The summary statistics (right) and stemplots (below) are based on natural logarithms; for common logarithms, multiply mean and standard deviation by 2.3026. For LPCB126, the zero terms were replaced $\ln 0.0026 \doteq -5.9522$, which accounts for the odd appearance of its stemplot.

	$\bar{x}$	s
LPCB28	−1.3345	1.1338
LPCB52	−0.7719	1.1891
LPCB118	0.8559	0.8272
LPCB126	−4.8457	0.7656
LPCB138	1.6139	0.8046
LPCB153	1.7034	0.9012
LPCB180	0.9752	0.9276
LPCB	3.9170	0.8020
LTEQ	0.8048	0.5966

LPCB28

−5	1
−4	
−4	
−3	6
−3	
−2	88765
−2	433331111100
−1	8888776665555
−1	443222211111
−0	99998888766555
−0	3100
0	4
0	56789
1	
1	9

LPCB52

−3	85
−3	
−2	776
−2	1100
−1	98877765
−1	4443211111000
−0	8777776666665
−0	443332111100
0	22334
0	5557
1	03
1	7
2	02

LPCB118

−1	4
−0	996
−0	33110
0	0112233334444444
0	556666788899999
1	012222222333334
1	556678899
2	111
2	59

LPCB126

−5	9999999999999999
−5	
−5	
−5	222
−5	111000
−4	99999999998888
−4	7776
−4	54444
−4	3322
−4	11100
−3	999888
−3	666
−3	554

LPCB138

−0	411
0	33
0	6677788999
1	0011112222333334
1	5555566777789
2	000001111222344
2	5555589
3	024

LPCB153

−0	110
0	134
0	5578899
1	0001122222222333444
1	5677888899999
2	01222233444
2	55569
3	01244
3	57

LPCB180

−0	975
−0	443110
0	00001111222334
0	567778889999
1	0011122222233344
1	55666689
2	012223
2	589
3	4

LPCB

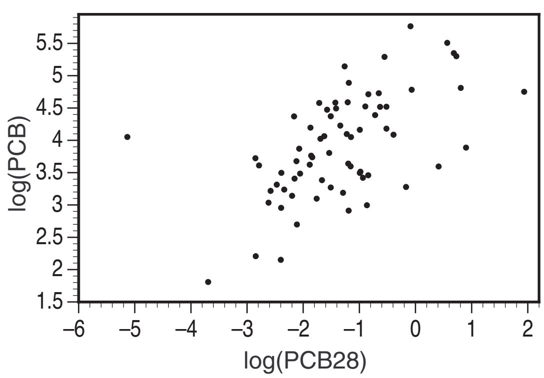
1	8
2	12
2	6999
3	0011222233444444
3	5556666777888
4	000000111233344
4	555555777788
5	1233
5	57

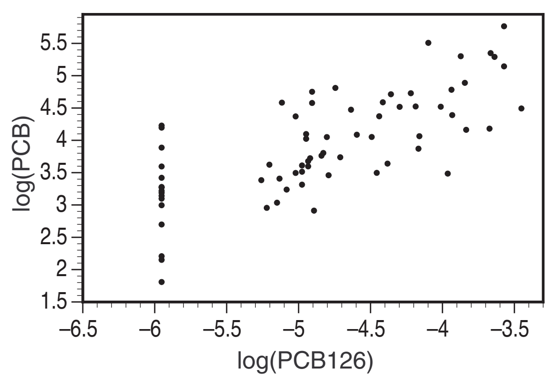
LTEQ

0	00000000111111
0	2223333333
0	4455555
0	66677
0	888899
1	0111111
1	22333
1	44455
1	6667777
1	888

11.47. **(a)** The correlations (all positive) are listed in the table below. The largest correlation is 0.956 (LPCB and LPCB138); the smallest (0.227, for LPCB28 and LPCB180) is not quite significantly different from 0 ($t = 1.91$, $P = 0.0607$) but, with 28 correlations, such a P-value could easily arise by chance, so we would not necessarily conclude that $\rho = 0$. Rather than showing all 28 scatterplots—which are all fairly linear and confirm the positive associations suggested by the correlations—we have included only two of the interesting ones: LPCB against LPCB28 and LPCB against LPCB126. The former is notable because of one outlier (specimen 39) in LPCB28; the latter stands out because of the "stack" of values in the LPCB126 data set that arose from the adjustment of the zero terms. (The outlier in LPCB28 and the stack in LPCB126 can be seen in other plots involving those variables; the two plots shown are the most appropriate for using the PCB congeners to predict LPCB, as the next exercise asks.) **(b)** All correlations are higher with the transformed data. In part, this is because these scatterplots do not exhibit the "greater scatter in the upper right" that was seen in many of the scatterplots of the original data.

	LPCB28	LPCB52	LPCB118	LPCB126	LPCB138	LPCB153	LPCB180
LPCB52	0.795						
LPCB118	0.533	0.671					
LPCB126	0.272	0.331	0.739				
LPCB138	0.387	0.540	0.890	0.792			
LPCB153	0.326	0.519	0.780	0.647	0.922		
LPCB180	0.227	0.301	0.654	0.695	0.896	0.867	
LPCB	0.570	0.701	0.906	0.729	0.956	0.905	0.829





11.48. Student results will vary with how many different models they try, and what tradeoff they consider between "good" (in terms of large R^2) and "simple" (in terms of the number of variables included in the model). The first Minitab output on the next page, produced with the `BREG` (best regression) command, gives some guidance as to likely answers; it shows the best models with one, two, three, four, five, six, and seven explanatory variables. We can see, for example, that if all variables are used, $R^2 = 0.975$, but we can achieve similar values of R^2 with fewer variables. The best regressions with two, three, and four explanatory variables are shown in the Minitab output on the next page.

Minitab output: Best subsets regression

```
                                        L L L L L
                                    L L P P P P P
                                    P P C C C C C
                                    C C B B B B B
                                    B B 1 1 1 1 1
              Adj.                  2 5 1 2 3 5 8
Vars   R-sq   R-sq    C-p        s  8 2 8 6 8 3 0

   1   91.4   91.3  141.4  0.23689          X
   2   96.2   96.1   28.5  0.15892    X     X
   3   96.8   96.6   16.1  0.14696  X X     X
   4   97.2   97.0    8.2  0.13826  X X     X X
   5   97.3   97.0    8.6  0.13776  X X X   X X
   6   97.5   97.2    6.0  0.13389  X X X   X X X
   7   97.5   97.2    8.0  0.13497  X X X X X X X
```

Best regression using two explanatory variables

```
LPCB = 2.74 + 0.175 LPCB52 + 0.813 LPCB138

Predictor       Coef       Stdev    t-ratio        p
Constant     2.74038     0.05860      46.76    0.000
LPCB52       0.17533     0.01926       9.10    0.000
LPCB138      0.81294     0.02846      28.56    0.000

s = 0.1589      R-sq = 96.2%     R-sq(adj) = 96.1%
```

Best regression using three explanatory variables

```
LPCB = 2.79 + 0.0908 LPCB28 + 0.104 LPCB52 + 0.821 LPCB138

Predictor       Coef       Stdev    t-ratio        p
Constant     2.79394     0.05633      49.60    0.000
LPCB28       0.09078     0.02601       3.49    0.001
LPCB52       0.10371     0.02717       3.82    0.000
LPCB138      0.82056     0.02641      31.07    0.000

s = 0.1470      R-sq = 96.8%     R-sq(adj) = 96.6%
```

Best regression using four explanatory variables

```
The regression equation is
LPCB = 2.79 + 0.107 LPCB28 + 0.0876 LPCB52 + 0.669 LPCB138 + 0.151 LPCB153

Predictor       Coef       Stdev    t-ratio        p
Constant     2.79081     0.05300      52.65    0.000
LPCB28       0.10684     0.02503       4.27    0.000
LPCB52       0.08763     0.02610       3.36    0.001
LPCB138      0.66854     0.05538      12.07    0.000
LPCB153      0.15118     0.04921       3.07    0.003

s = 0.1383      R-sq = 97.2%     R-sq(adj) = 97.0%
```

11.49. Using Minitab's `BREG` (best regression) command for guidance, we see that there is little improvement in R^2 beyond models with four explanatory variables. The best models with two, three, and four variables are given in the Minitab output below.

Minitab output: Best subsets regression

```
                                          L L L L L
                                      L L P P P P P
                                      P P C C C C C
                                      C C B B B B B
                                      B B 1 1 1 1 1
              Adj.                    2 5 1 2 3 5 8
Vars   R-sq   R-sq    C-p          s  8 2 8 6 8 3 0

   1   72.9   72.5   10.8   0.31266         X
   2   76.8   76.1    2.0   0.29166   X     X
   3   77.6   76.6    1.6   0.28859   X   X X
   4   78.0   76.7    2.5   0.28816   X   X X   X
   5   78.1   76.4    4.2   0.28981   X X X X   X
   6   78.2   76.1    6.1   0.29188   X X X X   X X
   7   78.2   75.7    8.0   0.29400   X X X X X X X
```

Best regression using two explanatory variables

```
The regression equation is
LTEQ = 3.96 + 0.107 LPCB28 + 0.622 LPCB126

Predictor        Coef       Stdev    t-ratio        p
Constant       3.9637      0.2275      17.42    0.000
LPCB28        0.10749     0.03242       3.32    0.001
LPCB126       0.62231     0.04801      12.96    0.000

s = 0.2917      R-sq = 76.8%     R-sq(adj) = 76.1%
```

Best regression using three explanatory variables

```
The regression equation is
LTEQ = 3.44 + 0.0777 LPCB28 + 0.114 LPCB118 + 0.543 LPCB126

Predictor        Coef       Stdev    t-ratio        p
Constant       3.4445      0.4029       8.55    0.000
LPCB28        0.07773     0.03736       2.08    0.041
LPCB118       0.11371     0.07319       1.55    0.125
LPCB126       0.54345     0.06952       7.82    0.000

s = 0.2886      R-sq = 77.6%     R-sq(adj) = 76.6%
```

Best regression using four explanatory variables

```
The regression equation is
LTEQ = 3.56 + 0.0720 LPCB28 + 0.170 LPCB118 + 0.554 LPCB126 - 0.0693 LPCB153

Predictor        Coef       Stdev    t-ratio        p
Constant       3.5568      0.4152       8.57    0.000
LPCB28        0.07199     0.03767       1.91    0.060
LPCB118       0.16973     0.08928       1.90    0.062
LPCB126       0.55374     0.07005       7.90    0.000
LPCB153      -0.06929     0.06344      -1.09    0.279

s = 0.2882      R-sq = 78.0%     R-sq(adj) = 76.7%
```

11.50. The degree of change in these elements of a regression can be readily seen by comparing the three regression results shown in the solution to Exercise 11.48; they will be even more visible if students have explored more models in their search for the best model. Student explanations might include observations of changes in particular coefficients from one model to another and perhaps might attempt to paraphrase the text's comments about *why* this happens.

11.51. In the table, two *IQR*s are given; those in parentheses are based on quartiles reported by Minitab, which computes quartiles in a slightly different way from this text's method.

	$\bar{x}$	M	s	IQR
Taste	24.53	20.95	16.26	23.9 (or 24.58)
Acetic	5.498	5.425	0.571	0.656 (or 0.713)
H2S	5.942	5.329	2.127	3.689 (or 3.766)
Lactic	1.442	1.450	0.3035	0.430 (or 0.4625)

None of the variables show striking deviations from Normality in the quantile plots (not shown). Taste and H2S are slightly right-skewed, and Acetic has two peaks. There are no outliers.

Taste

0	00
0	556
1	1234
1	55688
2	011
2	556
3	24
3	789
4	0
4	7
5	4
5	67

Acetic

4	455
4	67
4	8
5	1
5	2222333
5	444
5	677
5	888
6	0011
6	3
6	44

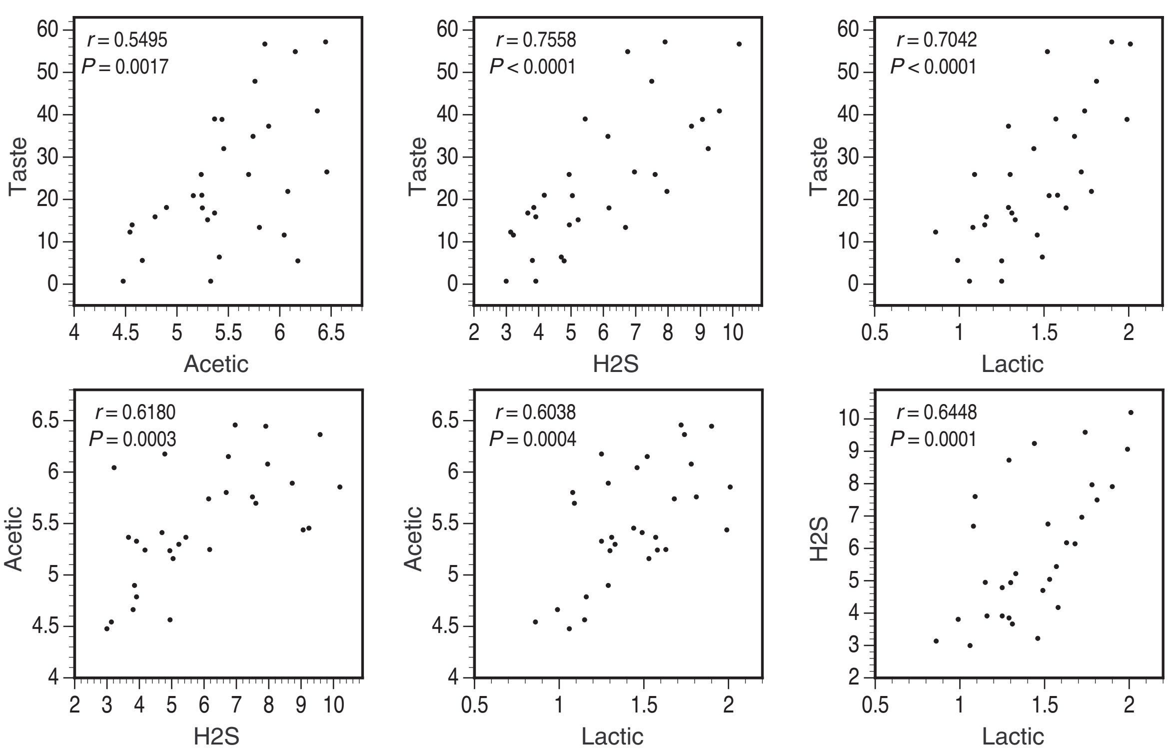
H2S

2	9
3	1268899
4	17799
5	024
6	11679
7	4699
8	7
9	025
10	1

Lactic

8	6
9	9
10	689
11	56
12	5599
13	013
14	469
15	2378
16	38
17	248
18	1
19	09
20	1

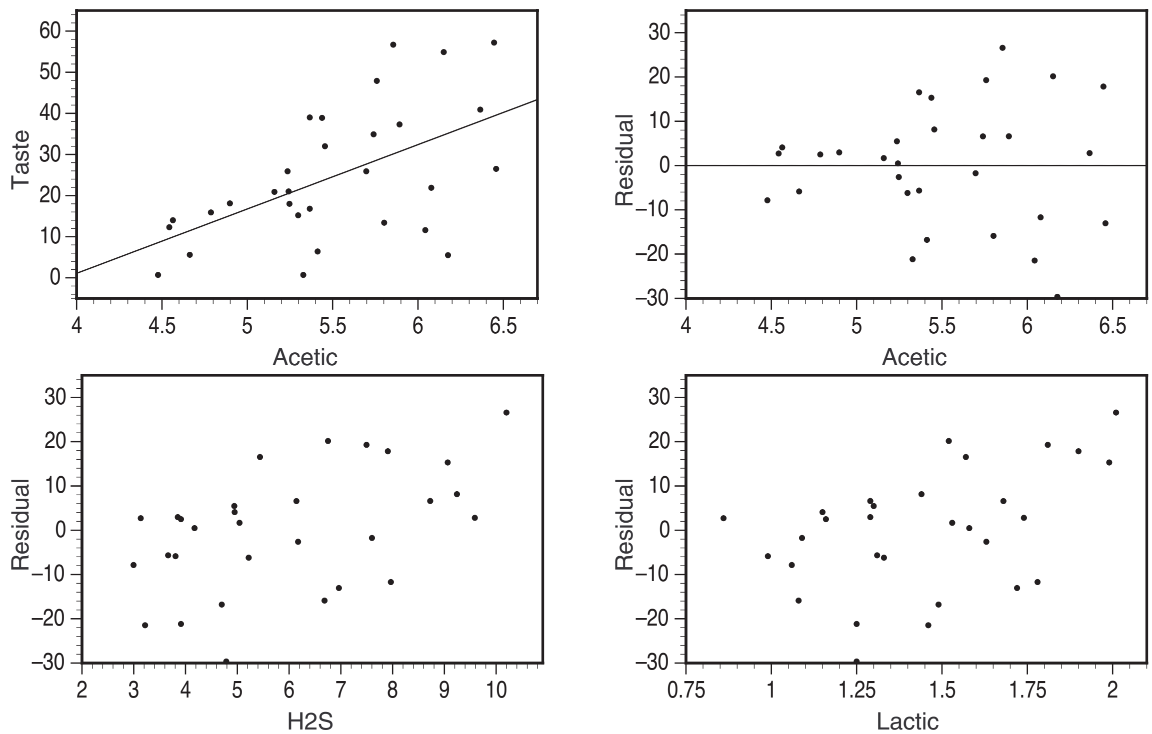
11.52. The plots show positive associations between the variables. The correlations and P-values are in the plots; all correlations are positive (as expected) and significantly different from 0. (Recall that the P-values are correct if the two variables are Normally distributed, in which case $t = r\sqrt{n-2}/\sqrt{1-r^2}$ has a $t(n-2)$ distribution if $\rho = 0$.)



11.53. The regression equation is $\widehat{\text{Taste}} = -61.50 + 15.648\ \text{Acetic}$ with $s = 13.82$ and $R^2 = 0.302$. The slope is significantly different from 0 ($t = 3.48$, $P = 0.002$).

Based on a stemplot (right) and quantile plot (not shown), the residuals seem to have a Normal distribution. Scatterplots (below) reveal positive associations between residuals and both H2S and Lactic. The plot of residuals against Acetic suggests greater scatter in the residuals for large Acetic values.

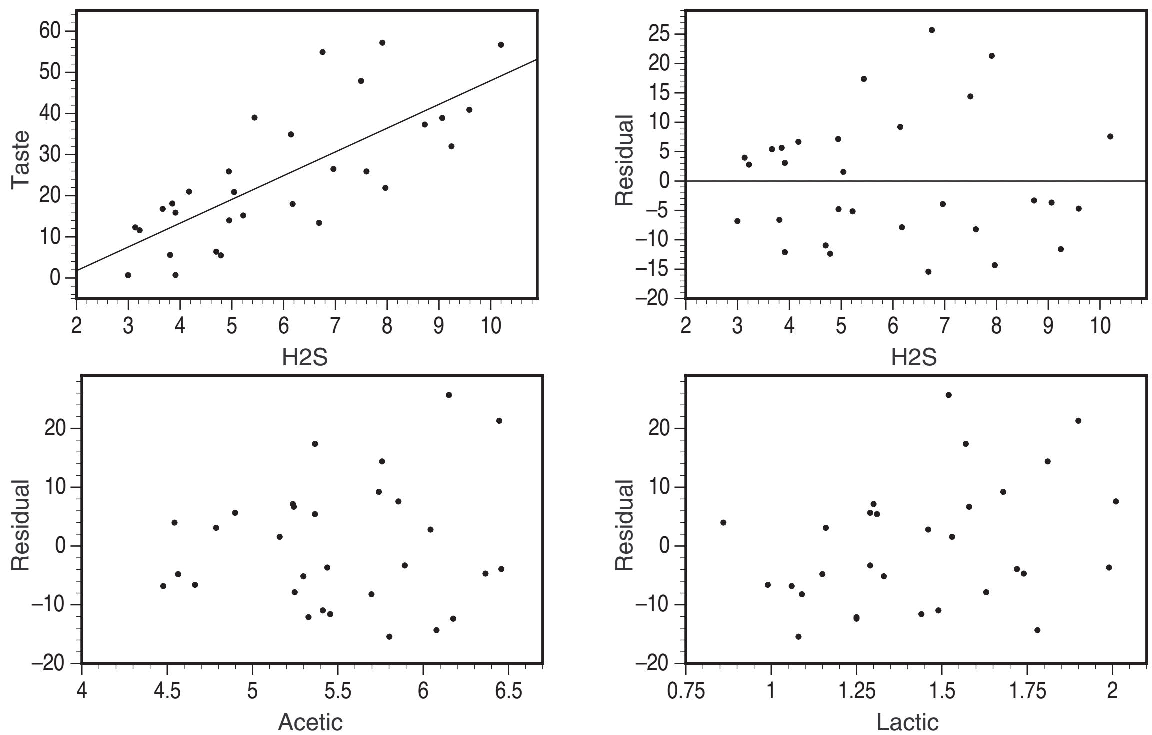
−2	9
−2	11
−1	65
−1	31
−0	7655
−0	21
0	0122224
0	5668
1	
1	5679
2	0
2	6



11.54. The regression equation is $\widehat{\text{Taste}} = -9.787 + 5.7761\,\text{H2S}$ with $s = 10.83$ and $R^2 = 0.571$. The slope is significantly different from 0 ($t = 6.11$, $P < 0.0005$).

Based on a stemplot (right) and quantile plot (not shown), the residuals may be slightly skewed, but do not differ greatly from a Normal distribution. Scatterplots (below) reveal weak positive associations between residuals and both Acetic and Lactic. The plot of residuals against H2S suggests greater scatter in the residuals for large H2S values.

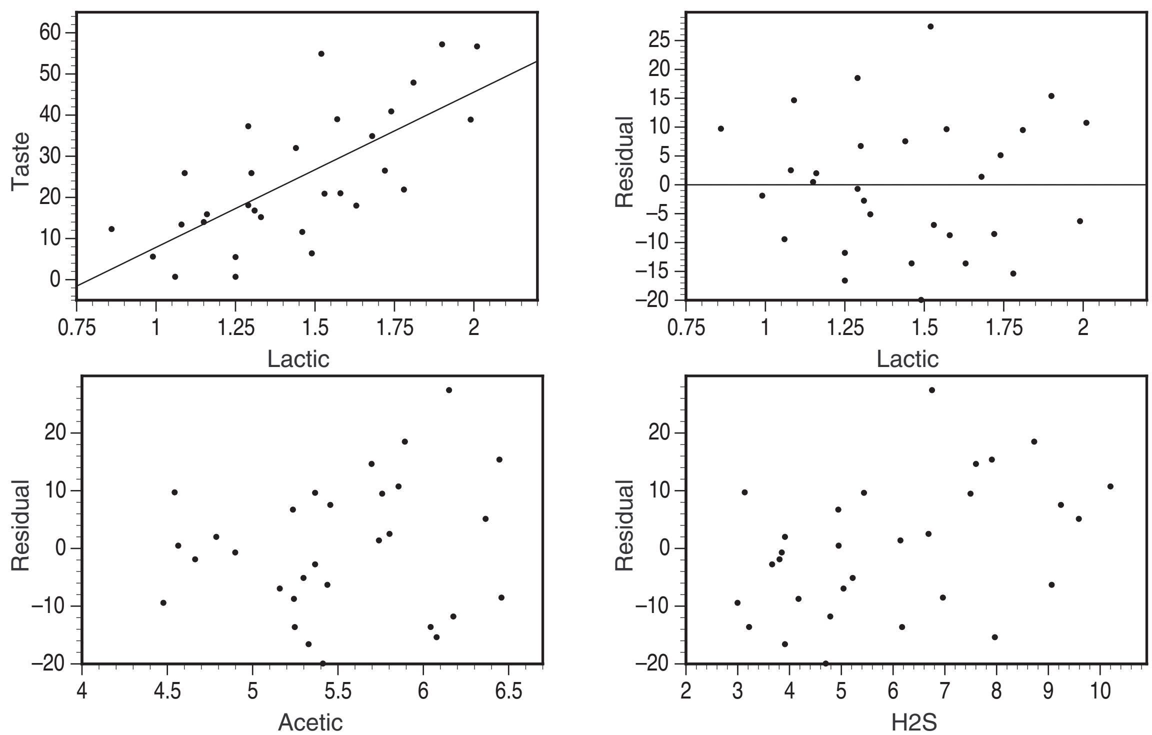
−1	5
−1	42210
−0	87665
−0	44333
0	1233
0	556779
1	4
1	7
2	1
2	5



11.55. The regression equation is $\widehat{\text{Taste}} = -29.86 + 37.720\ \text{Lactic}$ with $s = 11.75$ and $R^2 = 0.496$. The slope is significantly different from 0 ($t = 5.25$, $P < 0.0005$).

Based on a stemplot (right) and quantile plot (not shown), the residuals appear to be roughly Normal. Scatterplots (below) reveal no striking patterns for residuals vs. Acetic and H2S.

−1	965
−1	331
−0	988665
−0	210
0	0122
0	567999
1	04
1	58
2	
2	7



11.56. All information is in the table at the right. The intercepts differ from one model to the next because they represent different things—for example, in the first model, the intercept is the predicted value of Taste with Acetic = 0, etc.

x	$\widehat{\text{Taste}} =$	F	P	r^2	s
Acetic	$-61.50 + 15.648x$	12.11	0.002	30.2%	13.82
H2S	$-9.787 + 5.7761x$	37.29	<0.0005	57.1%	10.83
Lactic	$-29.86 + 37.720x$	27.55	<0.0005	49.6%	11.75

11.57. The regression equation is $\widehat{\text{Taste}} = -26.94 + 3.801\,\text{Acetic} + 5.146\,\text{H2S}$ with $s = 10.89$ and $R^2 = 0.582$. The t-value for the coefficient of Acetic is 0.84 ($P = 0.406$), indicating that it does not add significantly to the model when H2S is used because Acetic and H2S are correlated (in fact, $r = 0.618$ for these two variables). This model does a better job than any of the three simple linear regression models, but it is not much better than the model with H2S alone (which explained 57.1% of the variation in Taste)—as we might expect from the t-test result.

Minitab output: Regression of taste on acetic and h2s

```
taste = -26.9 + 3.80 acetic + 5.15 h2s

Predictor       Coef       Stdev    t-ratio        p
Constant      -26.94       21.19      -1.27    0.215
acetic         3.801       4.505       0.84    0.406
h2s            5.146       1.209       4.26    0.000

s = 10.89       R-sq = 58.2%     R-sq(adj) = 55.1%
```

11.58. The regression equation is $\widehat{\text{Taste}} = -27.592 + 3.946\,\text{H2S} + 19.887\,\text{Lactic}$ with $s = 9.942$. The model explains 65.2% of the variation in Taste, which is higher than for the two simple linear regressions. Both coefficients are significantly different from 0 ($P = 0.002$ for H2S, and $P = 0.019$ for Lactic).

Minitab output: Regression of taste on h2s and lactic

```
taste = -27.6 + 3.95 h2s + 19.9 lactic

Predictor       Coef       Stdev    t-ratio        p
Constant     -27.592       8.982      -3.07    0.005
h2s            3.946       1.136       3.47    0.002
lactic        19.887       7.959       2.50    0.019

s = 9.942       R-sq = 65.2%     R-sq(adj) = 62.6%
```

11.59. The regression equation is $\widehat{\text{Taste}} = -28.88 + 0.328\,\text{Acetic} + 3.912\,\text{H2S} + 19.671\,\text{Lactic}$ with $s = 10.13$. The model explains 65.2% of the variation in Taste (the same as for the model with only H2S and Lactic). Residuals of this regression appear to be Normally distributed and show no patterns in scatterplots with the explanatory variables. (These plots are not shown.)

The coefficient of Acetic is not significantly different from 0 ($P = 0.942$); there is no gain in adding Acetic to the model with H2S and Lactic. It appears that the best model is the H2S/Lactic model of Exercise 11.58.

Minitab output: Regression of taste on acetic, h2s, and lactic

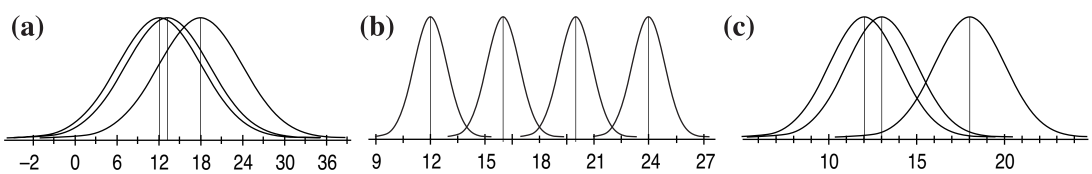
```
taste = -28.9 + 0.33 acetic + 3.91 h2s + 19.7 lactic

Predictor       Coef       Stdev    t-ratio        p
Constant      -28.88       19.74      -1.46    0.155
acetic         0.328       4.460       0.07    0.942
h2s            3.912       1.248       3.13    0.004
lactic        19.671       8.629       2.28    0.031

s = 10.13       R-sq = 65.2%     R-sq(adj) = 61.2%
```

Chapter 12 Solutions

12.1. **(a)** H_0 says the *population* means are all equal. **(b)** *Experiments* are best for establishing causation. **(c)** ANOVA is used to compare *means* (and assumes that the variances are equal). **(d)** Multiple comparisons procedures are used when we wish to determine which means are significantly different, but have no specific relations in mind before looking at the data. (Contrasts are used when we have prior expectations about the differences.)

12.2. **(a)** If we reject H_0, we conclude that *at least one* mean is different from the rest. **(b)** One-way ANOVA is used to compared two *or more* means. (When only means are to be compared, we usually use a two-sample t test.) **(c)** Two-way ANOVA is used to examine the effect of two *explanatory* variables (which have two *or more* values) on a response variable (which is assumed to have a Normal distribution, meaning that it can take any value, at least in theory).

12.3. We were given sample sizes $n_1 = 23$, $n_2 = 20$, and $n_3 = 28$ and standard deviations $s_1 = 5$, $s_2 = 5$, and $s_3 = 6$. **(a)** Yes: The guidelines for pooling standard deviations say that the ratio of largest to smallest should be less than 2; we have $\frac{6}{5} \doteq 1.2 < 2$. **(b)** Squaring the three standard deviations gives $s_1^2 = 25$, $s_2^2 = 25$, and $s_3^2 = 36$. **(c)** $s_p^2 = \dfrac{22s_1^2 + 19s_2^2 + 27s_3^2}{22 + 19 + 27} \doteq 29.3676$. **(d)** $s_p = \sqrt{s_p^2} \doteq 5.4192$.

12.4. 

12.5. **(a)** This sentence describes *between*-group variation. Within-group variation is the variation that occurs by chance among members of the same group. **(b)** The *sums of* squares (not the mean squares) in an ANOVA table will add. **(c)** The common population standard deviation σ (not its estimate s_p) is a parameter. **(d)** A small P means the means are not all the same, but the distributions may still overlap quite a bit. (See the "Caution" immediately preceding this exercise in the text.)

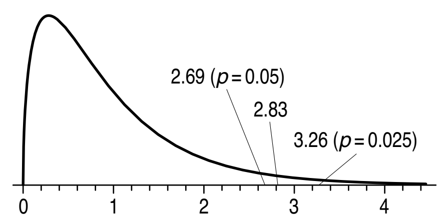
12.6. The answers are found in Table E (or using software) with $p = 0.05$ and degrees of freedom $I - 1$ and $N - I$. **(a)** $I = 4$, $N = 16$, df 3 and 12: $F > 3.49$ (software: 3.4903). **(b)** $I = 4$, $N = 24$, df 3 and 20: $F > 3.10$ (software: 3.0984). **(c)** $I = 4$, $N = 32$, df 3 and 28: $F > 2.95$ (software: 2.9467). **(d)** As the degrees of freedom increase, values from an F distribution tend to be smaller (closer to 1), so smaller values of F are statistically significant. In terms of ANOVA conclusions, we have learned that with smaller samples (fewer observations per group), the F statistic needs to be fairly large in order to reject H_0.

12.7. Assuming the t (ANOVA) test establishes that the means are different, contrasts and multiple comparison provide no further useful information. (With two means, there is only one comparison to make, and it has already been made by the t test.)

12.8. **(a)** The stated hypothesis is $\mu_{50\%} < \frac{1}{2}(\mu_{0\%} + \mu_{100\%})$, so we use the contrast $\psi = \frac{1}{2}(\mu_{0\%} + \mu_{100\%}) - \mu_{50\%}$, with coefficients 0.5, −1, and 0.5. The hypotheses can then be stated H_0: $\psi = 0$ versus H_a: $\psi > 0$. **(b)** The estimated contrast is $c = \frac{1}{2}(50 + 120) - 75 = 10\ \text{cm}^3$, with standard error $\text{SE}_c = s_p\sqrt{\frac{0.25}{40} + \frac{1}{40} + \frac{0.25}{40}} \doteq 5.8095$, so the test statistic is $t \doteq \frac{10}{5.8095} \doteq 1.7213$ with df $= 117$. The one-sided P-value is $P = 0.0439$, so this is significant at $\alpha = 0.05$, but not at $\alpha = 0.01$.

Note: *We wrote the contrast so that it would be positive when H_a is true (in keeping with the text's advice). We could also test this hypothesis using the contrast $\psi' = \mu_{50\%} - \frac{1}{2}(\mu_{0\%} + \mu_{100\%})$, or even $\psi'' = \mu_{0\%} + \mu_{100\%} - 2\mu_{50\%}$. The resulting t statistic is the same (except possibly in sign) regardless of the way the contrast is stated.*

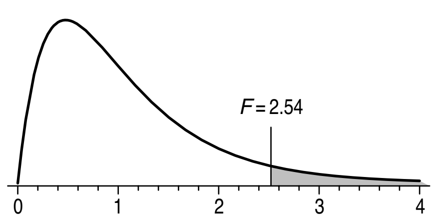
12.9. **(a)** With $I = 5$ groups and $N = 35$, we have df $I - 1 = 4$ and $N - I = 30$. In Table E, we see that $2.69 < F < 3.25$. **(b)** The sketch on the right shows the observed F value and the critical values from Table E. **(c)** $0.025 < P < 0.050$ (software gives 0.0420). **(d)** The alternative hypothesis states that at least one mean is different, not that all means are different.

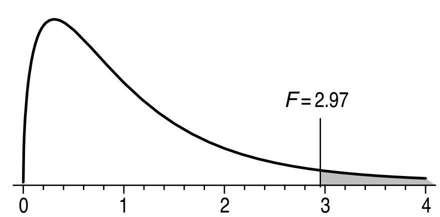


12.10. Compare each F statistic to an $F(I - 1, N - I)$ distribution:

	F	I	N	DFG	DFE	Critical values	P-value (Table E)	P-value (software)
(a)	2.69	7	35	6	28	$2.45 < F < 2.90$	$0.025 < P < 0.050$	0.0344
(b)	2.43	5	55	4	50	$2.06 < F < 2.56$	$0.050 < P < 0.100$	0.0597
(c)	3.06	6	34	5	28	$F = 3.06$	$P = 0.025$	0.0251

12.11. **(a)** $I = 3$ and $N = 33$, so the degrees of freedom are 2 and 30. $F = \frac{127}{50} = 2.54$. Comparing to the $F(2, 30)$ distribution in Table E, we find $2.49 < F < 3.32$, so $0.050 < P < 0.100$. (Software gives $P \doteq 0.0957$.) **(b)** $I = 4$ and $N = 32$, so the degrees of freedom are 3 and 28. $F = \frac{58/3}{182/28} \doteq 2.9744$. Comparing to the $F(3, 28)$ distribution in Table E, we find $2.95 < F < 3.63$, so $0.025 < P < 0.050$. (Software gives $P \doteq 0.0486$.)





12.12. **(a)** Yes: The guidelines for pooling standard deviations say that the ratio of largest to smallest should be less than 2; we have $\frac{42}{28} = 1.5$. **(b)** Squaring the three standard deviations gives $s_1^2 = 1369$, $s_2^2 = 784$, and $s_3^2 = 1764$. **(c)** $s_p^2 = \dfrac{28s_1^2 + 31s_2^2 + 120s_3^2}{28 + 31 + 120} \doteq 1532.49$. **(d)** $s_p = \sqrt{s_p^2} \doteq 39.15$. **(e)** Because the third sample was nearly twice as large as the other two put together, the pooled standard deviation is closest to s_3.

12.13. **(a)** Response: egg cholesterol level. Populations: chickens with different diets or drugs. $I = 3$, $n_1 = n_2 = n_3 = 25$, $N = 75$. **(b)** Response: rating on five-point scale. Populations: the three groups of students. $I = 3$, $n_1 = 31$, $n_2 = 18$, $n_3 = 45$, $N = 94$. **(c)** Response: quiz score. Populations: students in each TA group. $I = 3$, $n_1 = n_2 = n_3 = 14$, $N = 42$.

12.14. **(a)** Response: time to complete VR path. Populations: children using different navigation methods. $I = 4$, $n_i = 10$ $(i = 1, 2, 3, 4)$, $N = 40$. **(b)** Response: calcium content of bone. Populations: chicks eating diets with differing pesticide levels. $I = 5$, $n_i = 13$ $(i = 1, 2, 3, 4, 5)$, $N = 65$. **(c)** Response: total sales between 11:00 a.m. and 2:00 p.m. Populations: customers responding to one of four sample offers. $I = 4$, $n_i = 5$ $(i = 1, 2, 3, 4)$ and $N = 20$.

12.15. For all three situations, the hypotheses are H_0: $\mu_1 = \mu_2 = \mu_3$ versus H_a: at least one mean is different. The degrees of freedom are DFG = DFM = $I - 1$ (“model” or “between groups”), DFE = DFW = $N - I$ (“error” or “within groups”), and DFT = $N - 1$ (“total”). The degrees of freedom for the F test are DFG and DFE.

Situation	I	N	DFG	DFE	DFT	df for F statistic
Egg cholesterol level	3	75	2	72	74	$F(2, 72)$
Student opinions	3	94	2	91	93	$F(2, 91)$
Teaching assistants	3	42	2	39	41	$F(2, 39)$

12.16. For all three situations, the hypotheses are H_0: $\mu_1 = \mu_2 = \cdots = \mu_I$ versus H_a: at least one mean is different. The degrees of freedom are DFG = DFM = $I - 1$ (“model” or “between groups”), DFE = DFW = $N - I$ (“error” or “within groups”), and DFT = $N - 1$ (“total”). The degrees of freedom for the F test are DFG and DFE.

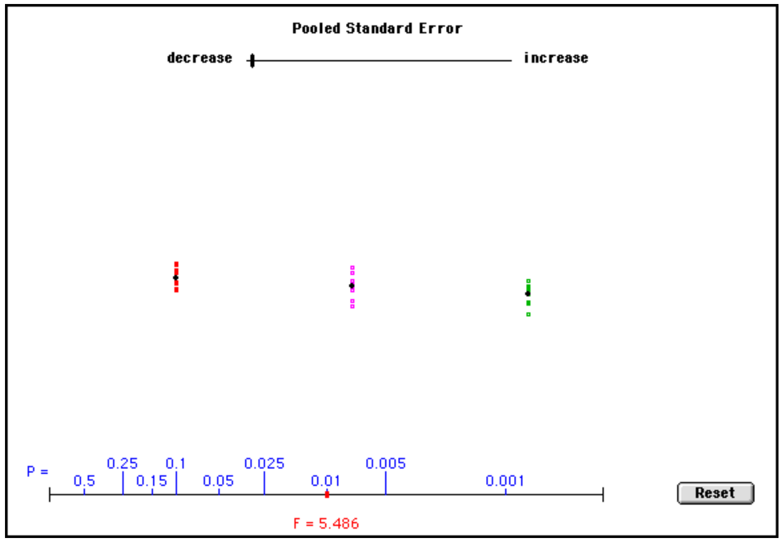
Situation	I	N	DFG	DFE	DFT	df for F statistic
VR navigation methods	4	40	3	36	39	$F(3, 36)$
Effect of pesticide on birds	5	65	4	60	64	$F(4, 60)$
Effect of free food on sales	4	20	3	16	19	$F(3, 16)$

12.17. **(a)** This sounds like a fairly well-designed experiment, so the results should at least apply to this farmer’s breed of chicken. **(b)** It would be good to know what proportion of the total student body falls in each of these groups—that is, is anyone overrepresented in this sample? **(c)** How well a TA teaches one topic (power calculations) might not reflect that TA’s overall effectiveness.

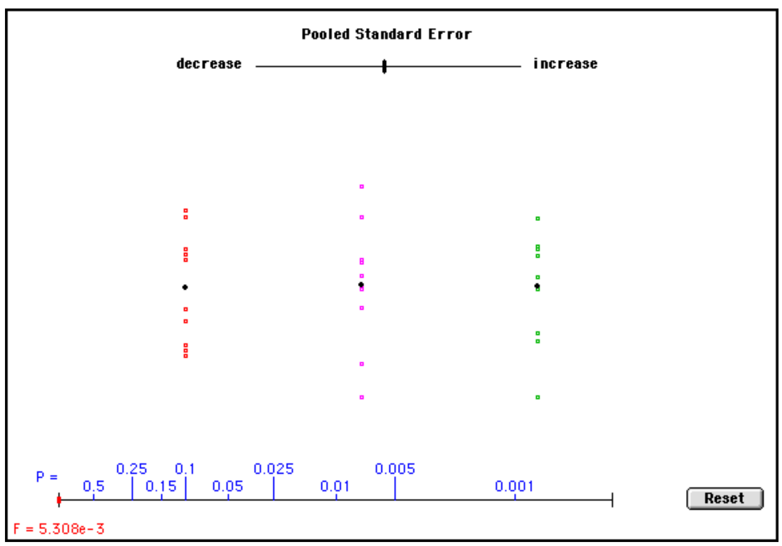
12.18. **(a)** This sounds like a fairly well-designed experiment, assuming the subjects come from a group which is representative of the population. (We assume that this teaching tool is intended for use with children and that the children used in the experiment were themselves deaf.) **(b)** This should at least give information about pesticide effect on bone calcium in chicks. It might not apply to adult chickens, or other species of birds. **(c)** The results *might* extend to similar sandwich shops, and perhaps to other times of day, or to weekend sales.

12.19. **(a)** With $I = 3$ and $N = 120$, we have df 2 and 117. **(b)** To use Table E, we compare to df 2 and 100; with $F > 5.02$, we conclude that $P < 0.001$. Software gives $P = 0.0003$. **(c)** Haggling and bargaining behavior is probably linked to the local culture, so we should hesitate to generalize these results beyond similar informal shops in Mexico.

12.20. **(a)** P-values close to 0.01 occur when F is close to 5.483. (This is the value for df 2 and 27; this applet seems to have three samples with 10 observations each.) How close students can get to this depends on how much they play around with the applet, and the pooled standard error setting. **(b)** As variation increases, F decreases and P increases.



12.21. **(a)** F can be made very small (close to 0), and P close to 1. **(b)** F increases, and P decreases. Moving the means farther apart means that (even with moderate spread) it is easier to see that the three groups represent three different populations (that is, populations having different means). Therefore, the evidence against H_0 becomes stronger.



12.22. We have $I = 4$ groups with $N = 620$. With the given group means, the overall mean is

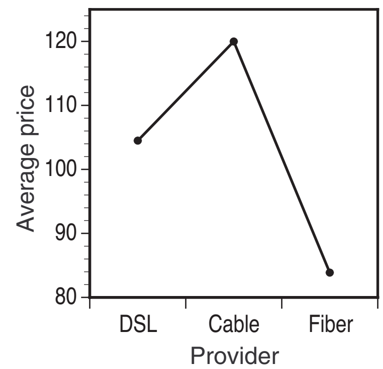
$$\bar{x} = \frac{130 \cdot 2.93 + 248 \cdot 3.00 + 174 \cdot 3.01 + 68 \cdot 3.39}{N} \doteq 3.0309$$

(a) DFG $= I - 1 = 3$ and DFE $= N - I = 616$. **(b)** The groups sum of squares is

$$\text{SSG} = 130(2.93 - \bar{x})^2 + 248(3.00 - \bar{x})^2 + 174(3.01 - \bar{x})^2 + 68(3.39 - \bar{x})^2 \doteq 10.4051$$

(c) $F = \dfrac{\text{MSG}}{\text{MSE}} = \dfrac{10.4051/3}{797.25/616} \doteq 2.68$. **(d)** Software gives $P \doteq 0.0461$, so we have enough evidence to reject H_0 at the 5% significance level. **(e)** The mean for the "other" group appears to be higher than the means of the first three groups (which are similar).

12.23. **(a)** Based on the sample means, fiber is cheapest and cable is most expensive. (Note that the providers are shown in this plot in the order given in the table, but they can be rearranged in any order.) **(b)** Yes; the smallest-to-largest standard deviation ratio is $\frac{40.39}{26.09} \doteq 1.55$. **(c)** The degrees of freedom are $I - 1 = 2$ and $N - I = 44$. From Table E (with df 2 and 40), we have $0.025 < P < 0.050$; software gives $P = 0.0427$. The difference in means is (barely) significant at the 5% level.



12.24. For the contrast $\psi = \frac{1}{2}(\mu_D + \mu_C) - \mu_F$, we test H_0: $\psi = 0$ versus H_a: $\psi > 0$. The pooled estimate of the standard deviation is

$$s_p = \sqrt{\frac{18s_1^2 + 19s_2^2 + 7s_3^2}{18 + 19 + 7}} \doteq 33.8170$$

The estimated contrast is $c = \frac{1}{2}(104.49 + 119.98) - 83.87 = \$28.365$, with standard error $\text{SE}_c = s_p\sqrt{\frac{0.25}{19} + \frac{0.25}{20} + \frac{1}{8}} \doteq 13.1259$, so the test statistic is $t = c/\text{SE}_c \doteq 2.161$ with df $= 44$. The one-sided P-value is $P = 0.0181$, so this is significant at $\alpha = 0.05$, but not at $\alpha = 0.01$.

12.25. **(a)** Use matched pairs t methods; we examine the change in reaction time for each subject. **(b)** No: We cannot use ANOVA methods because we do not have four independent samples. (The same group of subjects performed each of the four tasks.)

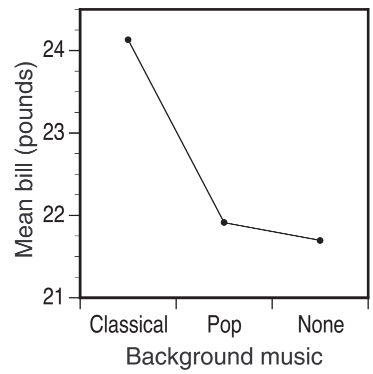
12.26. We have $\bar{x}_1 = 61.62$, $s_1 = 13.75$, $n_1 = 71$, $\bar{x}_2 = 46.47$, $s_2 = 7.94$, and $n_2 = 37$. For the pooled t procedure, we find $s_p \doteq 12.09$ and $t = 6.18$ (df $= 106$, $P < 0.0001$). The Minitab output below shows that $F = 38.17$ (t^2, up to rounding error).

Minitab output: ANOVA table

Source	DF	SS	MS	F	p
Factor	1	5583	5583	38.17	0.000
Error	106	15504	146		
Total	107	21087			

12.27. **(a)** With $I = 4$ and $N = 2290$, the degrees of freedom are DFG $= I - 1 = 3$ and DFE $= N - I = 2286$. **(b)** MSE $= s_p^2 = 4.6656$, so $F = \frac{\text{MSG}}{\text{MSE}} = \frac{11.806}{4.6656} \doteq 2.5304$. **(c)** The $F(3, 1000)$ entry in Table E gives $0.05 < P < 0.10$; software give $P \doteq 0.0555$.

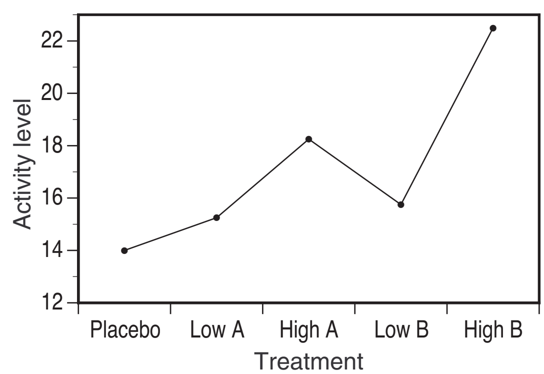
12.28. **(a)** The plot of means suggests that spending is higher for classical music, while pop and no music appear to have the same effect. **(b)** Yes: The guidelines for pooling standard deviations say that the ratio of largest to smallest should be less than 2; we have $\frac{3.332}{2.243} \doteq 1.49 < 2$. **(c)** The degrees of freedom are DFG $= I - 1 = 2$ and DFE $= N - I = 138$. Comparing to an $F(2, 100)$ distribution in Table E, we see that $P < 0.001$; software gives $P \doteq 0.00005$. We have strong evidence that the means are not all the same. **(d)** The higher average bill for classical music led to this conclusion; the difference between pop music and no background music is negligible. **(e)** The setting of this experiment ("a single high-end restaurant in England") might limit how much this conclusion can be generalized. It *might* extend to other high-end restaurants, but perhaps not to "family-style" restaurants, and almost certainly not to fast-food restaurants.



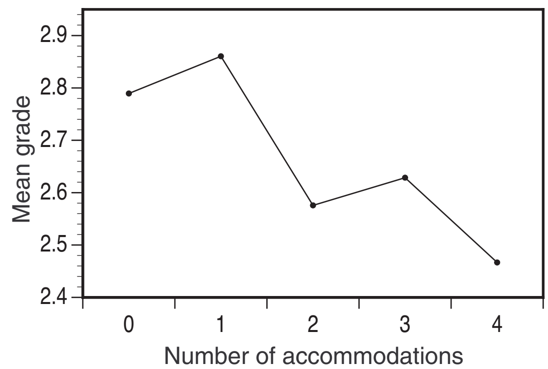
12.29. **(a)** The plot suggests that both drugs cause an increase in activity level, and Drug B appears to have a greater effect. **(b)** Yes: The guidelines for pooling standard deviations say that the ratio of largest to smallest should be less than 2; we have $\sqrt{\frac{14.00}{6.75}} \doteq 1.44 < 2$. The pooled variance is

$$s_p^2 = \frac{3(s_1^2 + s_2^2 + \cdots + s_5^2)}{3+3+3+3+3} = \frac{159}{15} = 10.6$$

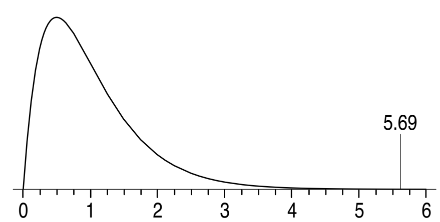
and $s_p = \sqrt{10.6} \doteq 3.2558$. **(c)** The degrees of freedom are DFG $= I - 1 = 4$ and DFE $= N - I = 15$. **(d)** Comparing to an $F(4, 15)$ distribution in Table E, we see that $3.80 < F < 4.89$, so $0.010 < P < 0.025$; software gives $P \doteq 0.0165$. We have significant evidence that the means are not all the same.



12.30. **(a)** It is useful to connect the points on the plot, to make the pattern (or lack thereof) more evident. There is some suggestion that average grade decreases as the number of accommodations increases. **(b)** Having too many decimal points is distracting; in this situation, no useful information is gained by having more than one or two digits after the decimal point. For example, the first mean and standard deviation would be more effectively presented as 2.79 and 0.85. **(c)** The largest-to-smallest SD ratio is slightly over 2 (about 2.009), so pooling is not advisable. (If we pool in spite of this, we find $s_p \doteq 0.8589$.) **(d)** Eliminating data points (without a legitimate reason) is always risky, although we could run the analysis with and without them. Combining the last three groups would be a bad idea if the data suggested that grades rebounded after 2 accommodations (i.e., if the average grades were higher for 3 and 4 accommodations), but as that is not the case, lumping 2, 3, and 4 accommodations seems reasonable. **(e)** ANOVA is not appropriate for these data, chiefly because we do not have 245 independent observations. **(f)** There may be a number of local factors (for example, student demographics or teachers' attitudes toward accommodations) which affected grades; these effects might not be the same elsewhere. **(g)** One weakness is that we do not have a control group for comparison; that is, we cannot tell what grades these students (or a similar group) would have had without accommodations.



12.31. **(a)** The variation in sample size is some cause for concern, but there can be no extreme outliers in a 1-to-7 scale, so ANOVA is probably reliable. **(b)** Pooling is reasonable: $\frac{1.26}{1.03} \doteq 1.22 < 2$. **(c)** With $I = 5$ groups and total sample size $N = 410$, we use an $F(4, 405)$ distribution. We can compare 5.69 to an $F(4, 200)$ distribution in Table E and conclude that $P < 0.001$, or with software determine that $P \doteq 0.0002$. **(d)** Hispanic Americans have the highest emotion scores, Japanese are in the middle, and the other three cultures are the lowest (and very similar).



12.32. **(a)** The largest-to-smallest SD ratios are 2.84, 1.23, and 1.14, so the text's guidelines are satisfied for intensity and recall, but not for frequency. **(b)** As in the previous exercise, $I = 5$ and $N = 410$, so we use an $F(4, 405)$ distribution. From the $F(4, 200)$ distribution in Table E, we can conclude that $P < 0.001$ for all three response variables. With software, we find that the P-values are much smaller; all are less than 0.00002. We conclude that, for each variable, we have strong evidence that some group mean is different. (This conclusion is cautious in the case of frequency because of our concern about the standard deviations.) **(c)** The table below shows one way of summarizing the means. For each variable, it attempts to identify low (underlined), medium, and high (boldface) values of that variable. Hispanic Americans were higher than other groups for all four variables. Asian Americans were low for all variables (the lowest in all but global score). Japanese were low on all but global score, while European Americans and Indians were in the middle for all but global score. **(d)** The results might not generalize to, for example, subjects who are from different parts of their countries or not in a college or university community. **(e)** Create a two-way table with counts of men and women in each cultural group. The Minitab output on the right gives $X^2 = 11.353$, df $= 4$, and $P = 0.023$, so we have evidence (significant at $\alpha = 0.05$) that the gender mix was not the same for all cultures. Specifically, Hispanic Americans and European Americans had higher percentages of women, which might further affect how much we can generalize the results.

Minitab output: Chi-square test

```
        Women     Men    Total
   1       38       8       46
        31.64   14.36

   2       22      11       33
        22.70   10.30

   3       57      34       91
        62.59   28.41

   4      102      58      160
       110.05   49.95

   5       63      17       80
        55.02   24.98

Total     282     128      410

ChiSq =  1.279 +  2.817 +
         0.021 +  0.047 +
         0.499 +  1.100 +
         0.589 +  1.297 +
         1.156 +  2.547 = 11.353
df = 4, p = 0.023
```

	Score	Frequency	Intensity	Recall
European Amer.	<u>4.39</u>	82.87	2.79	49.12
Asian Amer.	<u>4.35</u>	<u>72.68</u>	<u>2.37</u>	<u>39.77</u>
Japanese	4.72	<u>73.36</u>	<u>2.53</u>	<u>43.98</u>
Indian	<u>4.34</u>	82.71	2.87	49.86
Hispanic Amer.	**5.04**	**92.25**	**3.21**	**59.99**

12.33. Because the descriptions of these contrasts do not specify an expected direction for the comparison, the subtraction could be done either way (in the order shown, or in the opposite order). **(a)** $\psi_1 = \mu_2 - \frac{1}{2}(\mu_1 + \mu_4)$. **(b)** $\psi_2 = \frac{1}{3}(\mu_1 + \mu_2 + \mu_4) - \mu_3$.

12.34. Neither the descriptions in Exercise 12.33 nor the background information in Example 12.25 seem to give any indication of an expected direction for the contrasts, so we have given two-sided alternatives. If students give a one-sided alternative, they should explain why they did so. **(a)** With $\psi_1 = \mu_2 - \frac{1}{2}(\mu_1 + \mu_4)$ and $\psi_2 = \frac{1}{3}(\mu_1 + \mu_2 + \mu_4) - \mu_3$, we test H_0: $\psi_i = 0$ versus H_a: $\psi_i \neq 0$ (for $i = 1$ or 2). **(b)** The estimated contrasts are $c_1 \doteq 0.195$ and $c_2 \doteq 0.48$. **(c)** The pooled estimate of the standard deviation s_p is either 1.6771 or 1.6802 (see the note at the end of this solution), so $\text{SE}_{c_1} \doteq 0.3093$ or 0.3098, and $\text{SE}_{c_2} \doteq 0.2929$ or 0.2934. **(d)** Neither contrast is significantly different from 0 (with a two-sided alternative). For comparing brown eyes to the other colors, $t_1 \doteq 0.630$ or 0.629, with df $= 218$, for which $P \doteq 0.5290$ or 0.5298. For gaze up versus gaze down, $t_2 \doteq 1.639$ or 1.636, with df $= 218$, for which $P \doteq 0.1026$ or 0.1033. **(e)** The confidence intervals are $c_i \pm t^*\text{SE}_{c_i}$, where $t^* = 1.984$ (Table D) or 1.971 (software). This gives roughly -0.41 to 0.80 for ψ_1 and -0.10 to 1.06 for ψ_2.

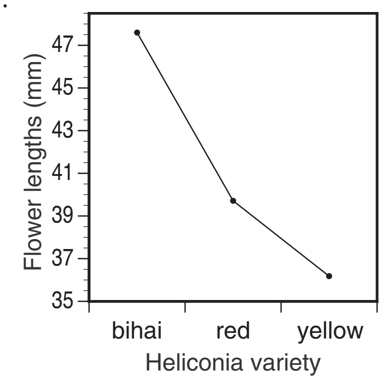
Note: *The simplest way to find the pooled standard deviation s_p is to use the value 1.6771 reported by SAS and Minitab in Figure 12.11 (or take $\sqrt{MSE}$ from the Excel output). Some students might compute it by hand from the numbers given in Example 12.25, which gives 1.6802. The difference is due to rounding; note that the reported standard deviation for brown eyes should be 1.72 rather than 1.73. In the end, our conclusions are the same either way.*

12.35. See the solution to Exercise 1.87 for stemplots. The means, standard deviations, and standard errors (all in millimeters) are given below. We reject H_0 and conclude that at least one mean is different ($F = 259.12$, df 2 and 51, $P < 0.0005$).

Variety	n	$\bar{x}$	s	$s_{\bar{x}}$
bihai	16	47.5975	1.2129	0.3032
red	23	39.7113	1.7988	0.3751
yellow	15	36.1800	0.9753	0.2518

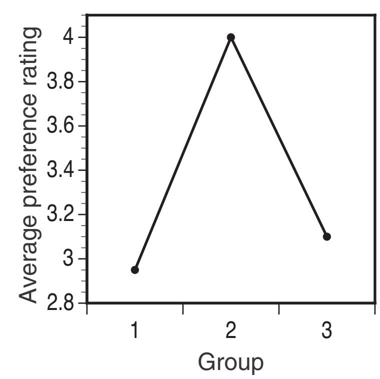
Minitab output: Analysis of Variance on length

```
Source    DF       SS       MS        F       p
Factor     2  1082.87   541.44   259.12   0.000
Error     51   106.57     2.09
Total     53  1189.44
```



12.36. **(a)** On the right are summary statistics and a plot of means; side-by-side stemplots are on the following page. Students might also use five-number summaries to describe the data, but with small samples and relatively unskewed distributions, they give us little additional information. **(b)** ANOVA gives $F \doteq 5.63$ (df 2 and 49) and $P \doteq 0.0063$—strong evidence of a difference in means. **(c)** Because preference ratings are whole numbers, the underlying distributions cannot be Normal, but apart from that, the stemplots and summary statistics show no particular causes for concern. On the following page are stemplots of the residuals, which show the expected granularity (due to the ratings being whole numbers). With such small samples, it is difficult to make any further judgments about Normality. **(d)** The three test statistics are

Group	n	$\bar{x}$	s
1	20	2.95	0.945
2	22	4.00	0.926
3	10	3.10	1.524



$$t_{12} = \frac{2.95 - 4.00}{s_p\sqrt{\frac{1}{20} + \frac{1}{22}}} \doteq -3.18, \quad t_{13} = \frac{2.95 - 3.10}{s_p\sqrt{\frac{1}{20} + \frac{1}{10}}} \doteq -0.36, \quad t_{23} = \frac{4.00 - 3.10}{s_p\sqrt{\frac{1}{22} + \frac{1}{10}}} \doteq 2.21$$

Results will vary with the method used, and the overall significance level. Using the Bonferroni method with $\alpha = 0.05$ (and three comparisons), we have $t^{**} \doteq 2.479$, so only groups 1 and 2 are significantly different.

Minitab output: Analysis of Variance on Preference

```
Source     DF      SS      MS      F      p
Factor      2   12.84    6.42   5.63  0.006
Error      49   55.85    1.14
Total      51   68.69
```

```
Group 1            Group 2            Group 3
 1 | 00             1 |                1 | 0
 2 | 00             2 | 00             2 | 000
 3 | 000000000000   3 | 000            3 | 000
 4 | 000            4 | 0000000000     4 | 0
 5 | 0              5 | 0000000        5 | 0
 6 |                6 |                6 | 0
```

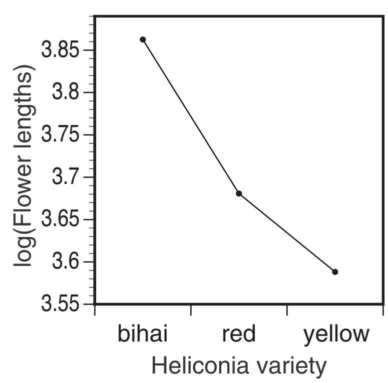
```
Group 1 residuals    Group 2 residuals    Group 3 residuals    All residuals
−2 |                 −2 | 00              −2 | 1               −2 | 100
−1 | 99              −1 | 000             −1 | 111             −1 | 99111000
−0 | 99              −0 | 00000           −0 | 111             −0 | 9911100000
 0 | 000000000000     0 | 00000            0 | 9                0 | 000000000000000009
 1 | 000              1 | 0000000          1 | 9                1 | 00000000009
 2 | 0                2 |                  2 | 9                2 | 09
```

12.37. Stemplots are shown on the right; means, standard deviations, and standard errors are given below. We reject H_0 and conclude that at least one mean is different ($F = 244.27$, df 2 and 51, $P < 0.0005$). These results are essentially the same as in Exercise 12.35.

```
bihai              red               yellow
38 | 33            36 | 233333       35 | 44
38 | 44444555      36 | 4445         35 | 667
38 | 7777          36 | 667          35 | 8889
38 |               36 | 8            36 | 00011
39 | 11            37 | 00           36 |
                   37 | 23333        36 | 4
                   37 | 4
                   37 | 6
```

Note: *All of the numbers in these samples are between 34 and 50; for that range of input values, the log function closely resembles a line. Because a linear transformation has no effect on skewness, the effect of this transformation is minimal, as can be confirmed by comparing these stemplots to those in the solution to Exercise 1.87, and this means plot with the one in the solution to Exercise 12.35.*

Variety	n	$\bar{x}$	s	$s_{\bar{x}}$
bihai	16	3.8625	0.02515	0.006286
red	23	3.6807	0.04496	0.009374
yellow	15	3.5882	0.02698	0.006966



Minitab output: Analysis of Variance on log-length

```
Source     DF        SS        MS       F      p
Factor      2   0.61438   0.30719  244.27  0.000
Error      51   0.06414   0.00126
Total      53   0.67852
```

12.38. **(a)** Statistics and plots are below. **(b)** The standard deviations satisfy the text's guidelines for pooling. One concern is that all three distributions are slightly left-skewed and the youngest nonfiction death is an outlier. **(c)** ANOVA gives $F = 6.56$ (df 2 and 120) and $P = 0.002$, so we conclude that at least one mean is different. **(d)** The appropriate contrast is $\psi_1 = \frac{1}{2}(\mu_{\text{nov}} + \mu_{\text{nf}}) - \mu_{\text{p}}$. (This is defined so that the $\psi_1 > 0$ if poets die younger. This is not absolutely necessary but is in keeping with the text's advice.) The null hypothesis is H_0: $\psi_1 = 0$; the Yeats quote hardly seems like an adequate reason to choose a one-sided alternative, but students may have other opinions. For the test, we compute $c \doteq 10.9739$, $\text{SE}_c \doteq 3.0808$, and $t \doteq 3.56$ with df $= 120$. The P-value is very small regardless of whether H_a is one- or two-sided, so we conclude that the contrast is positive (and poets die young). **(e)** For this comparison, the contrast is $\psi_2 = \mu_{\text{nov}} - \mu_{\text{nf}}$, and the hypotheses are H_0: $\psi_2 = 0$ versus H_a: $\psi_2 \neq 0$. (Because the alternative is two-sided, the subtraction in this contrast can go either way.) For the test, we compute $c \doteq -5.4272$, $\text{SE}_c \doteq 3.4397$, and $t \doteq -1.58$ with df $= 120$. This gives $P = 0.1172$; the difference between novelists and nonfiction writers is not significant. **(f)** With three comparisons and df $= 120$, the Bonferroni critical value is $t^{**} = 2.4280$. The pooled standard deviation is $s_p \doteq 14.4592$, so the differences, standard errors, and t values are:

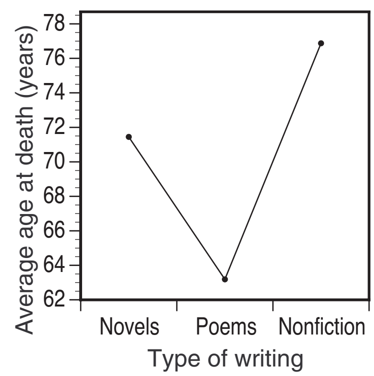
$$\begin{aligned}
\bar{x}_{\text{nov}} - \bar{x}_{\text{p}} &\doteq 8.2603, & \text{SE}_{\text{nov}-\text{p}} &= s_p\sqrt{\tfrac{1}{67} + \tfrac{1}{32}} \doteq 3.1071, & t &\doteq 2.66 \\
\bar{x}_{\text{nov}} - \bar{x}_{\text{nf}} &\doteq -5.4272, & \text{SE}_{\text{nov}-\text{nf}} &= s_p\sqrt{\tfrac{1}{67} + \tfrac{1}{24}} \doteq 3.4397, & t &\doteq -1.58 \\
\bar{x}_{\text{p}} - \bar{x}_{\text{nf}} &\doteq -13.6875, & \text{SE}_{\text{p}-\text{nf}} &= s_p\sqrt{\tfrac{1}{32} + \tfrac{1}{24}} \doteq 3.9044, & t &\doteq -3.51
\end{aligned}$$

The first and last differences are greater (in absolute value) than t^{**}, so those differences are significant. The second difference is the same one tested in the contrast of part (e); the standard error and the conclusion are the same.

	n	$\bar{x}$	s	$s_{\bar{x}}$
Novels	67	71.4478	13.0515	1.5945
Poems	32	63.1875	17.2971	3.0577
Nonfiction	24	76.8750	14.0969	2.8775

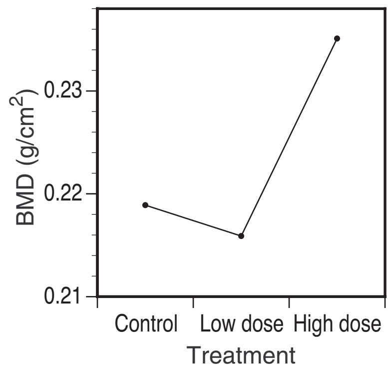
Minitab output: Analysis of Variance on age at death

```
Source    DF       SS       MS       F       p
Writer     2     2744     1372    6.56   0.002
Error    120    25088      209
Total    122    27832
```



12.39. **(a)** The means, standard deviations, and standard errors are given below (all in grams per cm^2). **(b)** All three distributions appear to reasonably close to Normal, and the standard deviations are suitable for pooling. **(c)** ANOVA gives $F = 7.72$ (df 2 and 42) and $P = 0.001$, so we conclude that the means are not all the same. **(d)** With df $= 42$, 3 comparisons, and $\alpha = 0.05$, the Bonferroni critical value is $t^{**} = 2.4937$. The pooled standard deviation is $s_p \doteq 0.01437$ and the standard error of each difference is $\text{SE}_D = s_p\sqrt{1/15 + 1/15} \doteq 0.005246$, so two means are significantly different if they differ by $t^{**}\text{SE}_D \doteq 0.01308$. The high-dose mean is significantly different from the other two. **(e)** Briefly: High doses of kudzu isoflavones increase BMD.

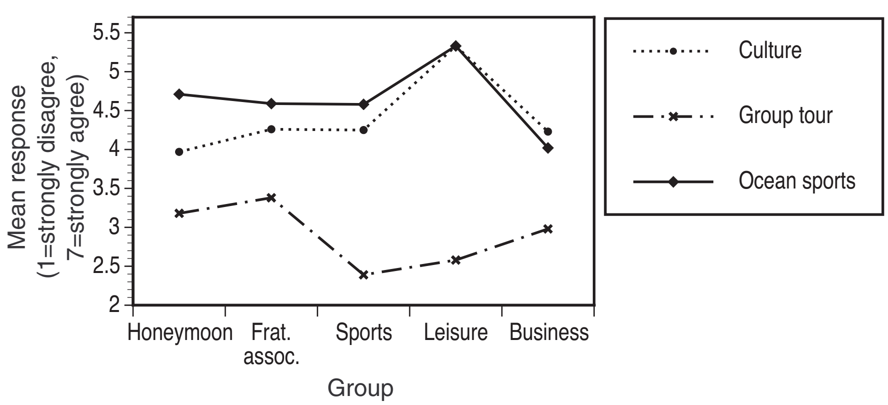
	n	$\bar{x}$	s	$s_{\bar{x}}$
Control	15	0.2189	0.01159	0.002992
Low dose	15	0.2159	0.01151	0.002972
High dose	15	0.2351	0.01877	0.004847



Minitab output: Analysis of Variance on BMD

```
Source     DF        SS        MS        F       p
Factor      2  0.003186  0.001593     7.72   0.001
Error      42  0.008668  0.000206
Total      44  0.011853
```

12.40. **(a)** With $I = 5$ and $N = 315$, the F tests have df 4 and 310. **(b)** The response variable does not need to be Normally distributed; rather, the deviations from the mean within each group should be Normal. **(c)** By comparing each F statistic to 2.401—the 5% critical value for an $F(4, 310)$ distribution—we see that the means are significantly different for the first three questions. (We can also compute the P-value for each F statistic to reach this conclusion.) **(d)** A possible plot is shown below; this could also be split into three separate plots. For the first question, the leisure group appears to have the most interest in experiencing Hawaiian culture. For the second question, the sports and leisure groups had a lower preference for a group tour, while fraternal associations had a higher preference. For the third question, the leisure group is most interested in ocean sports, and the business group is least interested.



12.41. (a) The mean responses were not significantly different for the last question. **(b)** Taking the square roots of the given values of MSE gives the values of s_p. For the Bonferroni method with $\alpha = 0.05$, df = 310, and 10 comparisons, $t^{**} = 2.827$. Only the largest difference within each set of means is significant:

$$t_{14} = \frac{3.97 - 5.33}{1.8058\sqrt{\frac{1}{34} + \frac{1}{26}}} \doteq -2.891 \quad \text{experience culture—honeymoon and leisure groups}$$

$$t_{23} = \frac{3.38 - 2.39}{1.6855\sqrt{\frac{1}{56} + \frac{1}{105}}} \doteq 3.550 \quad \text{group tour—fraternal association/sports groups}$$

$$t_{45} = \frac{5.33 - 4.02}{2.0700\sqrt{\frac{1}{26} + \frac{1}{94}}} \doteq 2.856 \quad \text{ocean sports—leisure/business groups}$$

12.42. (a) At right. **(b)** We test H_0: $\mu_1 = \cdots = \mu_4$ versus H_a: not all μ_i are equal. ANOVA gives $F = 9.24$ with df 3 and 74, for which $P < 0.0005$, so we reject the null hypothesis. The type of lesson does affect the mean score change; in particular, it appears that students who take piano lessons had significantly higher scores than the other students.

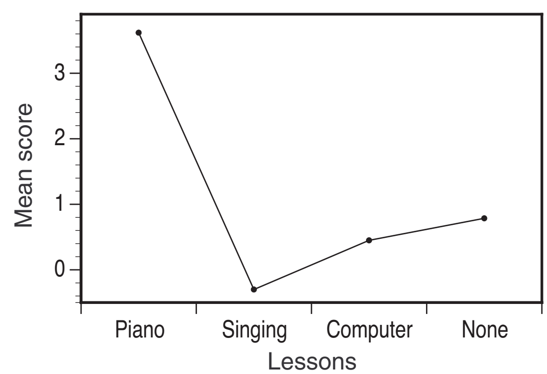
Lesson	n	$\bar{x}$	s	$s_{\bar{x}}$
Piano	34	3.6176	3.0552	0.5240
Singing	10	−0.3000	1.4944	0.4726
Computer	20	0.4500	2.2118	0.4946
None	14	0.7857	3.1908	0.8528

Minitab output: Analysis of Variance on scores

```
Source     DF        SS        MS        F        p
Lesson      3    207.28     69.09     9.24    0.000
Error      74    553.44      7.48
Total      77    760.72
```

12.43. We have six comparisons to make, and df = 74, so the Bonferroni critical value with $\alpha = 0.05$ is $t^{**} = 2.7111$. The pooled standard deviation is $s_p \doteq 2.7348$. The table below shows the differences, their standard errors, and the t statistics.

The Piano mean is significantly higher than the other three, but the other three means are not significantly different.

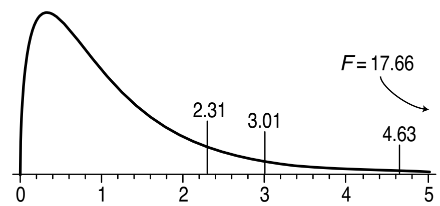


$D_{PS} =$ 3.91765	$D_{PC} =$ 3.16765	$D_{PN} =$ 2.83193
$SE_{PS} =$ 0.98380	$SE_{PC} =$ 0.77066	$SE_{PN} =$ 0.86843
$t_{PS} =$ 3.982	$t_{PC} =$ 4.110	$t_{PN} =$ 3.261
	$D_{SC} = -0.75000$	$D_{SN} = -1.08571$
	$SE_{SC} =$ 1.05917	$SE_{SN} =$ 1.13230
	$t_{SC} = -0.708$	$t_{SN} = -0.959$
		$D_{CN} = -0.33571$
		$SE_{CN} =$ 0.95297
		$t_{CN} = -0.352$

12.44. We test the hypothesis H_0: $\psi = \mu_P - \frac{1}{3}(\mu_S + \mu_C + \mu_N) = 0$; the sample contrast is $c \doteq 3.618 - \frac{1}{3}(-0.300 + 0.450 + 0.786) = 3.306$. The pooled standard deviation estimate is $s_p = 2.735$, so $SE_c = 2.735\sqrt{1/34 + \frac{1}{9}/10 + \frac{1}{9}/20 + \frac{1}{9}/14} \doteq 0.6356$. Then $t = 3.306/0.6356 \doteq 5.20$, with df $= 74$. This is enough evidence ($P < 0.001$) to reject H_0 in favor of H_a: $\psi > 0$, so we conclude that mean score changes for piano students are greater than the average of the means for the other three groups.

12.45. **(a)** Pooling is reasonable: The ratio is $\frac{0.824}{0.657} \doteq 1.25$. For the pooled standard deviation, we compute

$$s_p^2 = \frac{488s_1^2 + 68s_2^2 + 211s_3^2}{488 + 68 + 211} \doteq 0.5902$$

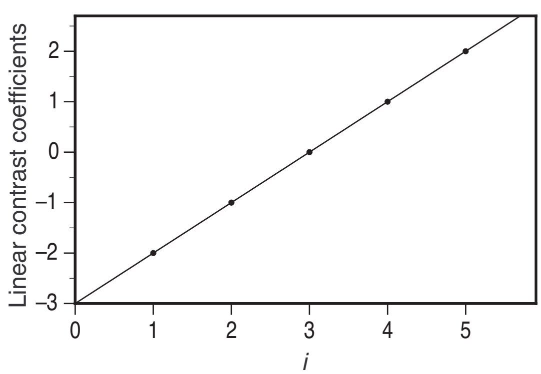


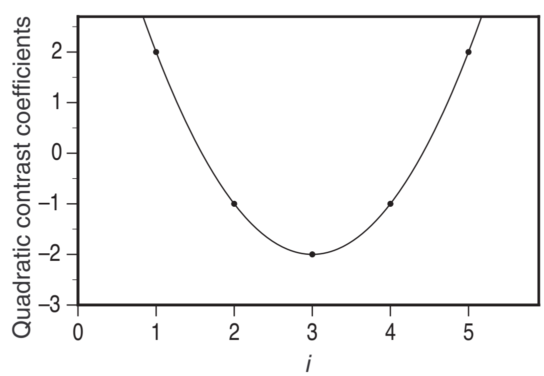
so $s_p \doteq \sqrt{0.5902} \doteq 0.7683$. **(b)** Comparing $F = 17.66$ to an $F(2, 767)$ distribution, we find $P < 0.001$. Sketches of this distribution will vary; in the graph on the right, the three marked points are the 10%, 5%, and 1% critical values, so we can see that the observed value lies well above the bulk of this distribution. **(c)** For the contrast $\psi = \mu_2 - \frac{1}{2}(\mu_1 + \mu_3)$, we test H_0: $\psi = 0$ versus H_a: $\psi > 0$. We find $c \doteq 0.585$ with $SE_c \doteq 0.0977$, so $t = c/SE_c \doteq 5.99$ with df $= 767$, and $P < 0.0001$.

12.46. **(a)** The linear-trend coefficients (plot below, left) fall on a line. If $\mu_1 = \mu_2 = \cdots = \mu_5$, then the linear-trend contrast $\psi_1 = -2\mu_1 - 1\mu_2 + 0\mu_3 + 1\mu_4 + 2\mu_5 = 0$. **(b)** The quadratic-trend coefficients (plot below, right) fall on a parabola. If all μ_i are equal, then $\psi_2 = 2\mu_1 - 1\mu_2 - 2\mu_3 - 1\mu_4 + 2\mu_5 = 0$. If $\mu_i = 5i$, then

$$\psi_2 = 2 \cdot 5 - 1 \cdot 10 - 2 \cdot 15 - 1 \cdot 20 + 2 \cdot 25 = 0$$

(c) The sample contrasts are $c_2 \doteq -3.36$ and $c_3 \doteq 1.12$. **(d)** The standard errors are $SE_{c_2} \doteq 0.8306$ and $SE_{c_3} \doteq 0.6425$, so the test statistics are $t_2 \doteq -3.364$ and $t_3 \doteq 1.740$. With df $= 129$, the P-values are 0.001 and 0.0842. Combined with the linear-trend result from Example 12.20 ($t = -0.18$, $P = 0.861$), we see that we have significant evidence for a quadratic trend, but not for a linear or cubic trend.





12.47. **(a)** Pooling is reasonable, as the largest-to-smallest ratio is about 1.65. **(b)** ANOVA gives $F = 7.98$ (df 2 and 27), for which $P = 0.002$. We reject H_0 and conclude that not all means are equal.

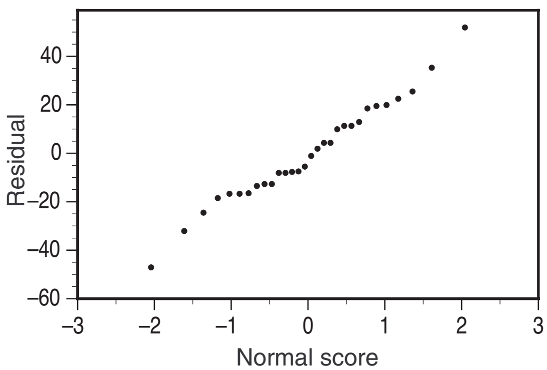
	n	$\bar{x}$	s
Control	10	601.1	27.364
Low jump	10	612.5	19.329
High jump	10	638.7	16.594

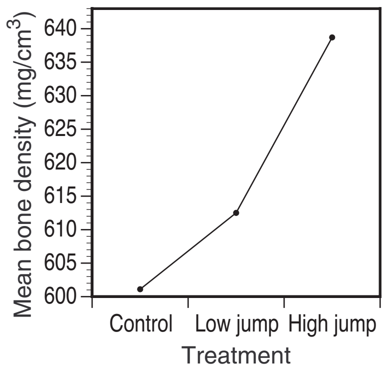
Minitab output: Analysis of Variance on density

```
Source     DF      SS      MS       F       p
Treatment   2    7434    3717    7.98   0.002
Error      27   12580     466
Total      29   20013
```

12.48. **(a)** The residuals appear to be reasonably Normal. **(b)** With df = 27, three comparisons, and $\alpha = 0.05$, the Bonferroni critical value is $t^{**} = 2.5525$. The pooled standard deviation is $s_p \doteq 21.5849$, and the standard error of each difference is $\text{SE}_D = s_p\sqrt{1/10 + 1/10} \doteq 9.6531$, so two means are significantly different if they differ by $t^{**}\text{SE}_D \doteq 24.6390$. The high-jump mean is significantly different from the other two.

```
−4 | 7
−3 | 2
−2 | 4
−1 | 8666322
−0 | 887751
 0 | 1449
 1 | 112899
 2 | 25
 3 | 5
 4 |
 5 | 1
```





12.49. **(a)** Pooling is risky because $\frac{0.6283}{0.2520} = 2.49 > 2$. **(b)** ANOVA gives $F = 31.16$ (df 2 and 9), for which $P < 0.0005$. We reject H_0 and conclude that not all means are equal.

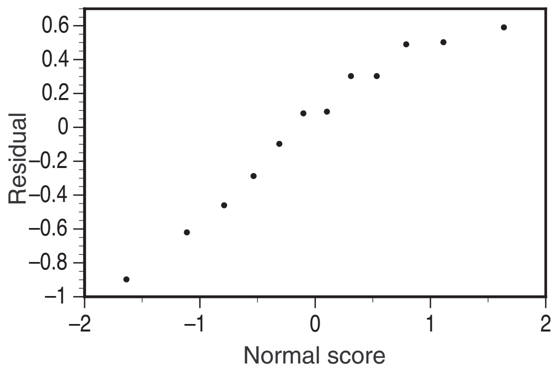
	n	$\bar{x}$	s
Aluminum	4	2.0575	0.2520
Clay	4	2.1775	0.6213
Iron	4	4.6800	0.6283

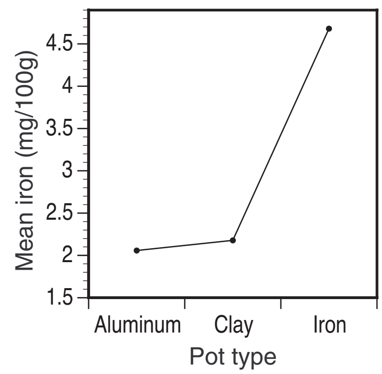
Minitab output: Analysis of Variance on iron

```
Source     DF      SS      MS       F       p
Pot         2  17.539   8.770   31.16   0.000
Error       9   2.533   0.281
Total      11  20.072
```

12.50. **(a)** There are no clear violations of Normality, but the number of residuals is so small that it is difficult to draw any conclusions. **(b)** With df = 9, three comparisons, and $\alpha = 0.05$, the Bonferroni critical value is $t^{**} = 2.9333$. The pooled standard deviation is $s_p \doteq 0.5305$, and the standard error of each difference is $\text{SE}_D = s_p\sqrt{1/4 + 1/4} \doteq 0.3751$, so two means are significantly different if they differ by $t^{**}\text{SE}_D \doteq 1.1003$. The iron mean is significantly higher than the other two.

```
-0 | 8
-0 | 6
-0 | 4
-0 | 2
-0 | 0
 0 | 00
 0 | 33
 0 | 455
```





12.51. **(a)** Pooling is risky because $\frac{8.66}{2.89} = 3 > 2$.
(b) ANOVA gives $F = 137.94$ (df 5 and 12), for which $P < 0.0005$. We reject H_0 and conclude that not all means are equal.

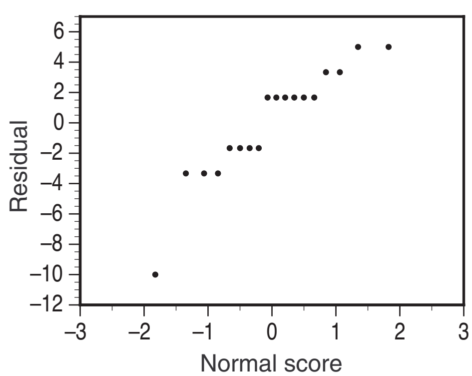
	n	$\bar{x}$	s
ECM1	3	65.0%	8.6603%
ECM2	3	$63.\overline{3}$%	2.8868%
ECM3	3	$73.\overline{3}$%	2.8868%
MAT1	3	$23.\overline{3}$%	2.8868%
MAT2	3	$6.\overline{6}$%	2.8868%
MAT3	3	$11.\overline{6}$%	2.8868%

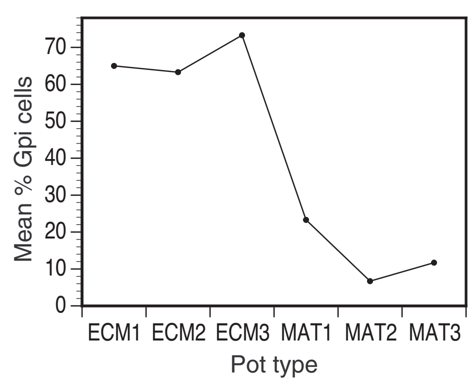
Minitab output: Analysis of Variance on Gpi

```
Source     DF       SS        MS        F        p
Treatment   5   13411.1    2682.2   137.94    0.000
Error      12     233.3      19.4
Total      17   13644.4
```

12.52. **(a)** The residuals have one low outlier, and a lot of granularity, so Normality is difficult to assess. **(b)** With df = 12, 15 comparisons, and $\alpha = 0.05$, the Bonferroni critical value is $t^{**} = 3.6489$. The pooled standard deviation is $s_p \doteq 4.4096\%$, and the standard error of each difference is $\text{SE}_D = s_p\sqrt{1/3 + 1/3} \doteq 3.6004\%$, so two means are significantly different if they differ by $t^{**}\text{SE}_D \doteq 13.1375\%$. The three ECM means are significantly higher than the three MAT means. **(c)** The contrast is $\psi = \frac{1}{3}(\mu_{\text{ECM1}} + \mu_{\text{ECM2}} + \mu_{\text{ECM3}}) - \frac{1}{3}(\mu_{\text{MAT1}} + \mu_{\text{MAT2}} + \mu_{\text{MAT3}})$, and the hypotheses are H_0: $\psi = 0$ versus H_a: $\psi \neq 0$. For the test, we compute $c \doteq 53.33\%$, $\text{SE}_c \doteq 2.0787\%$, and $t \doteq 25.66$ with df = 12. This has a tiny P-value; the difference between ECM and MAT is highly significant. This is consistent with the Bonferroni results from part (b).

```
-1 | 0
-0 |
-0 |
-0 |
-0 | 333
-0 | 1111
 0 | 111111
 0 | 33
 0 | 55
```





12.53. Let μ_1 be the placebo mean, μ_2 and μ_3 be the means for low and high doses of Drug A, and μ_4 and μ_5 be the means for low and high doses of Drug B. Recall that $s_p \doteq 3.2558$. **(a)** The first contrast is $\psi_1 = \mu_1 - \frac{1}{2}(\mu_2 + \mu_4)$; the second is $\psi_2 = \mu_3 - \mu_2 - (\mu_5 - \mu_4)$. **(b)** The estimated contrasts are $c_1 = 14.00 - 0.5(15.25) - 0.5(15.75) = -1.5$ and $c_2 = (18.25 - 15.25) - (22.50 - 15.75) = -3.75$. The respective standard errors are:

$$\mathrm{SE}_{c_1} = s_p\sqrt{\frac{1}{4} + \frac{0.25}{4} + \frac{0}{4} + \frac{0.25}{4} + \frac{0}{4}} \doteq 1.9937 \text{ and}$$

$$\mathrm{SE}_{c_2} = s_p\sqrt{\frac{0}{4} + \frac{1}{4} + \frac{1}{4} + \frac{1}{4} + \frac{1}{4}} = s_p \doteq 3.2558$$

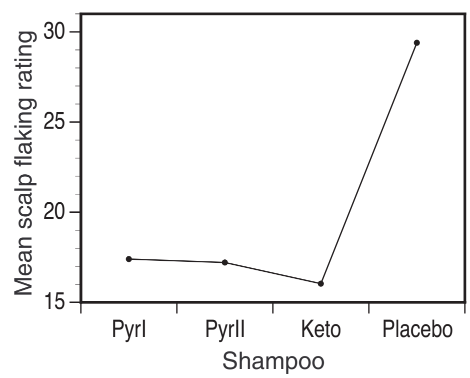
(c) Neither contrast is significant ($t_1 \doteq -0.752$ and $t_2 \doteq -1.152$, for which the one-sided P-values are 0.2317 and 0.1337). We do not have enough evidence to conclude that low doses increase activity level over a placebo, nor can we conclude that activity level changes due to increased dosage are different between the two drugs.

12.54. **(a)** Below. **(b)** To test H_0: $\mu_1 = \cdots = \mu_4$ versus H_a: not all μ_i are equal, ANOVA (Minitab output below) gives $F = 967.82$ (df 3 and 351), which has $P < 0.0005$. We conclude that not all means are equal; specifically, the "Placebo" mean is much higher than the other three means.

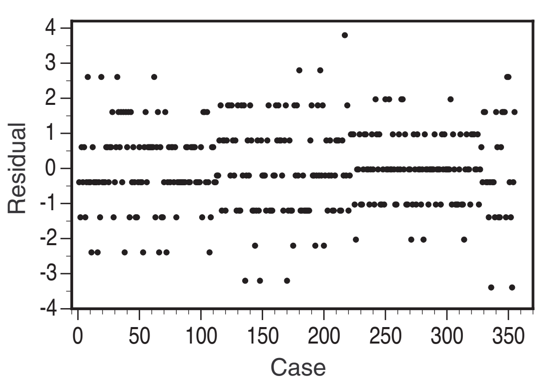
Shampoo	n	$\bar{x}$	s	$s_{\bar{x}}$
PyrI	112	17.393	1.142	0.108
PyrII	109	17.202	1.352	0.130
Keto	106	16.028	0.931	0.090
Placebo	28	29.393	1.595	0.301

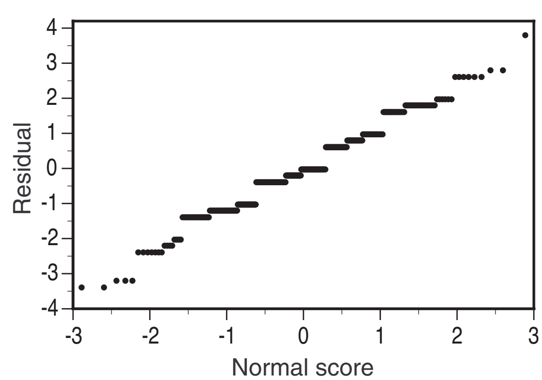
Minitab output: Analysis of Variance on Flaking

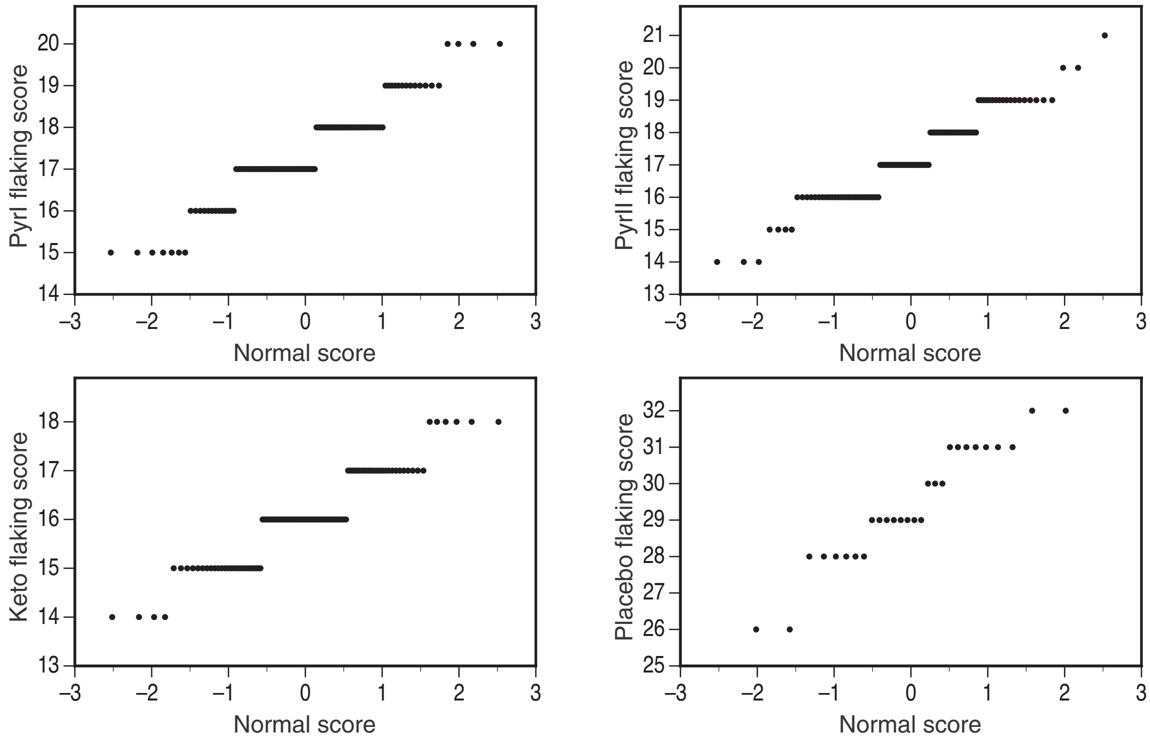
```
Source    DF       SS       MS       F       p
Code       3  4151.43  1383.81  967.82   0.000
Error    351   501.87     1.43
Total    354  4653.30
```



12.55. **(a)** The plot (below) shows granularity (which varies between groups), but that should not make us question independence; it is due to the fact that the scores are all integers. **(b)** The ratio of the largest to the smallest standard deviations is $1.595/0.931 \doteq 1.714$—less than 2. **(c)** Apart from the granularity, the quantile plots (on the following page) are reasonably straight. **(d)** Again, apart from the granularity, the residual quantile plot (below, right) looks pretty good.







12.56. We have six comparisons to make, and df = 351, so the Bonferroni critical value with $\alpha = 0.05$ is $t^{**} = 2.6533$. The pooled standard deviation is $s_p = \sqrt{\text{MSE}} \doteq 1.1958$; the differences, standard errors, and t statistics are below. The only *non*significant difference is between the two Pyr treatments (meaning the second application of the shampoo is of little benefit). The Keto shampoo mean is the lowest; the placebo mean is by far the highest.

$D_{\text{Py1-Py2}} = 0.19102$	$D_{\text{Py1-K}} = 1.36456$	$D_{\text{Py1-P}} = -12.0000$
$\text{SE}_{\text{Py1-Py2}} = 0.16088$	$\text{SE}_{\text{Py1-K}} = 0.16203$	$\text{SE}_{\text{Py1-P}} = 0.25265$
$t_{\text{Py1-Py2}} = 1.187$	$t_{\text{Py1-K}} = 8.421$	$t_{\text{Py1-P}} = -47.497$
	$D_{\text{Py2-K}} = 1.17353$	$D_{\text{Py2-P}} = -12.1910$
	$\text{SE}_{\text{Py2-K}} = 0.16312$	$\text{SE}_{\text{Py2-P}} = 0.25334$
	$t_{\text{Py2-K}} = 7.195$	$t_{\text{Py2-P}} = -48.121$
		$D_{\text{K-P}} = -13.3646$
		$\text{SE}_{\text{K-P}} = 0.25407$
		$t_{\text{K-P}} = -52.601$

12.57. **(a)** The three contrasts are:

$$\psi_1 = \tfrac{1}{3}\mu_{\text{Py1}} + \tfrac{1}{3}\mu_{\text{Py2}} + \tfrac{1}{3}\mu_{\text{K}} - \mu_{\text{P}}$$
$$\psi_2 = \tfrac{1}{2}\mu_{\text{Py1}} + \tfrac{1}{2}\mu_{\text{Py2}} - \mu_{\text{K}}$$
$$\psi_3 = \mu_{\text{Py1}} - \mu_{\text{Py2}}$$

$c_1 = -12.51$	$c_2 = 1.269$	$c_3 = 0.191$
$\text{SE}_{c_1} \doteq 0.2355$	$\text{SE}_{c_2} \doteq 0.1413$	$\text{SE}_{c_3} \doteq 0.1609$
$t_1 = -53.17$	$t_2 = 8.98$	$t_3 = 1.19$
$P_1 < 0.0005$	$P_2 < 0.0005$	$P_3 \doteq 0.2359$

(b) The pooled standard deviation is $s_p = \sqrt{\text{MSE}} \doteq 1.1958$. The estimated contrasts and their standard errors are in the table. For example:

$$\text{SE}_{c_1} = s_p\sqrt{\tfrac{1}{9}/112 + \tfrac{1}{9}/109 + \tfrac{1}{9}/106 + 1/28} \doteq 0.2355$$

(c) We test H_0: $\psi_i = 0$ versus H_a: $\psi_i \neq 0$ for each contrast. The t- and P-values are given in the table. The Placebo mean is significantly higher than the average of the other three, while the Keto mean is significantly lower than the average of the two Pyr means. The difference between the Pyr means is not significant (meaning the second application of the shampoo is of little benefit)—this agrees with our conclusion from Exercise 12.56.

12.58. **(a)** At right. **(b)** Each new value (except for n) is simply (old value)/100. (Standard errors were not computed for Exercise 12.51, but for all groups, we simply divide by $\sqrt{3}$.) **(c)** The SS and MS entries differ from those of Exercise 12.51—by a factor of $0.0001 = (1/100)^2$. However, everything else is the same: $F = 137.94$ with df 5 and 12; $P < 0.0005$, so we (again) reject H_0 and conclude that not all means are equal.

	n	$\bar{x}$	s	$s_{\bar{x}}$
ECM1	3	0.65	0.08660	0.05
ECM2	3	$0.63\overline{3}$	0.02887	0.01667
ECM3	3	$0.73\overline{3}$	0.02887	0.01667
MAT1	3	$0.23\overline{3}$	0.02887	0.01667
MAT2	3	$0.06\overline{6}$	0.02887	0.01667
MAT3	3	$0.11\overline{6}$	0.02887	0.01667

Minitab output: Analysis of Variance on GpiPct

```
Source      DF       SS        MS        F        p
Treatment    5   1.34111   0.26822   137.94    0.000
Error       12   0.02333   0.00194
Total       17   1.36444
```

12.59. Because the measurements in Exercise 12.51 are percents, the instructions to "add 5% to each response" could be interpreted in two ways:

(1) new response = old response + 5

(2) new response = old response × 1.05

The table on the right gives summary statistics for both interpretations (all numbers in percents). For (1), the means increase by 5, but everything else remains the same; the ANOVA table is identical to the one in the solution to Exercise 12.51. For (2), both the means and standard deviations are multiplied by 1.05, SS and MS are multiplied by 1.05^2, but F and P remain the same (ANOVA table below).

	Version (1)		Version (2)	
	$\bar{x}$	s	$\bar{x}$	s
ECM1	70.0	8.6603	68.25	9.0933
ECM2	$68.\overline{3}$	2.8868	66.5	3.0311
ECM3	$78.\overline{3}$	2.8868	77	3.0311
MAT1	$28.\overline{3}$	2.8868	24.5	3.0311
MAT2	$11.\overline{6}$	2.8868	7	3.0311
MAT3	$16.\overline{6}$	2.8868	12.25	3.0311

Minitab output: Analysis of Variance on GpiVers2

```
Source      DF       SS        MS        F        p
Treatment    5   14785.8    2957.1   137.94    0.000
Error       12     257.3      21.4
Total       17   15043.0
```

12.60. There is no effect on the test statistic, df, P-value, and conclusion. The degrees of freedom are not affected, because the number of groups and sample sizes are unchanged; meanwhile, the SS and MS values change (by a factor of b^2), but this change does not affect F because the factors of b^2 cancel out in the ratio $F = \text{MSG/MSE}$. With the same F- and df values, the P-value and conclusion are necessarily unchanged.

Proof of these statements is not too difficult, but it requires careful use of the SS formulas. For most students, a demonstration with several choices of a and b would probably be more convincing than a proof. However, here is the basic idea: Using results of Chapter 1, we know that the means undergo the same transformation as the data ($\bar{x}_i^* = a + b\bar{x}_i$), while the standard deviations are changed by a factor of $|b|$. Let $\bar{x}$ be the average of *all* the data; note that $\bar{x}^* = a + b\bar{x}$. Now $\text{SSG} = \sum_{i=1}^{I} n_i(\bar{x}_i - \bar{x})^2$, so:

$$\text{SSG}^* = \sum_i n_i(\bar{x}_i^* - \bar{x}^*)^2 = \sum_i n_i(b\,\bar{x}_i - b\,\bar{x})^2 = \sum_i n_i\, b^2(\bar{x}_i - \bar{x})^2 = b^2\text{SSG}$$

Similarly, we can establish that $\text{SSE}^* = b^2\text{SSE}$ and $\text{SST}^* = b^2\text{SST}$. Since the MS values are merely SS values divided by the (unchanged) degrees of freedom, these also change by a factor of b^2.

12.61. A table of means and standard deviations is below. Quantile plots are not shown, but apart from the granularity of the scores and a few possible outliers, there are no marked deviations from Normality. Pooling is reasonable for both PRE1 and PRE2; the ratios are 1.24 and 1.48.

For both PRE1 and PRE2, we test H_0: $\mu_B = \mu_D = \mu_S$ versus H_a: at least one mean is different. Both tests have df 2 and 63. For PRE1, $F = 1.13$ and $P = 0.329$; for PRE2, $F = 0.11$ and $P = 0.895$. There is no reason to believe that the mean pretest scores differ between methods.

		PRE1		PRE2	
Method	n	$\bar{x}$	s	$\bar{x}$	s
Basal	22	10.5	2.9721	$5.\overline{27}$	2.7634
DRTA	22	$9.\overline{72}$	2.6936	$5.\overline{09}$	1.9978
Strat	22	$9.1\overline{36}$	3.3423	$4.9\overline{54}$	1.8639

Minitab output: Analysis of Variance on PRE1

```
Source     DF       SS       MS       F       p
Group       2    20.58    10.29    1.13   0.329
Error      63   572.45     9.09
Total      65   593.03
```

Analysis of Variance on PRE2

```
Source     DF       SS       MS       F       p
Group       2     1.12     0.56    0.11   0.895
Error      63   317.14     5.03
Total      65   318.26
```

12.62. Stemplots and summary statistics are on the following page. Some of the distributions have mild outliers or skewness, but there are no serious violations of Normality evident. Pooling is appropriate for all three response variables.

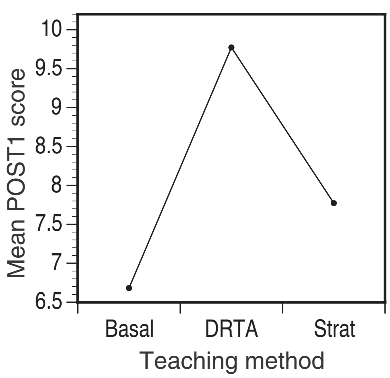
Variable	s_p	SE_D	$t^{**}\mathrm{SE}_D$
POST1	3.1885	0.9614	2.3646
POST2	2.3785	0.7171	1.7639
POST3	6.3141	1.9038	4.6825

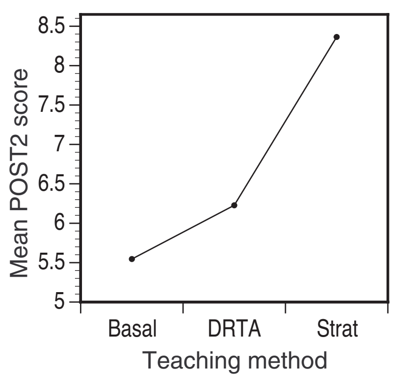
The three F statistics, all with df 2 and 63, are 5.32 ($P = 0.007$), 8.41 ($P = 0.001$), and 4.48 ($P = 0.015$). We conclude that at least one mean is different for each posttest.

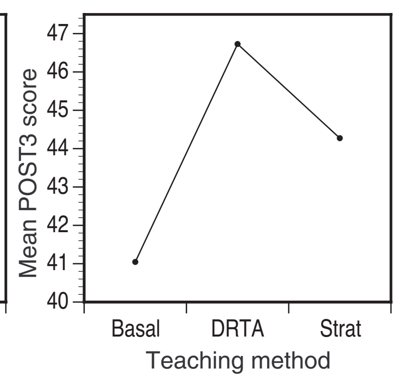
For multiple comparisons, we have three comparisons, df 63, and $\alpha = 0.05$, so the Bonferroni critical value is $t^{**} = 2.4596$. In the table (above, right) are the pooled standard deviations, standard error of each difference, and values of $t^{**}\mathrm{SE}_D$ (the "minimum significant difference," or MSD) for each response variable. For both POST1 and POST3, DRTA is significantly greater than Basal, but no other comparisons are significant. For POST2, Strat is significantly greater than both Basal and DRTA.

We also may examine the contrasts $\psi_1 = -\mu_\mathrm{B} + \frac{1}{2}(\mu_\mathrm{D} + \mu_\mathrm{S})$, which is positive if the average of the new methods is greater than the basal mean, and $\psi_2 = \mu_\mathrm{D} - \mu_\mathrm{S}$, which compares the two new methods. Estimated contrasts, standard errors, t, and P are given in the table below. (The P-values are one-sided for ψ_1 and two-sided for ψ_2.) We see that c_1 is significantly positive for all three variables. The second contrast was included in our multiple comparisons, where we found the difference significant only for POST2, but this time, when we are testing *only* this difference, rather than all three possible differences, we conclude that $\mu_\mathrm{D} > \mu_\mathrm{S}$ for POST1, in addition to the difference for POST2.

Variable	c_1	SE_{c_1}	t_1	P_1	c_2	SE_{c_2}	t_2	P_2
POST1	2.0909	0.8326	2.51	0.0073	2.0000	0.9614	2.08	0.0415
POST2	1.7500	0.6211	2.82	0.0032	−2.1364	0.7171	−2.98	0.0041
POST3	4.4545	1.6487	2.70	0.0044	2.4545	1.9038	1.29	0.2019







Test	Method	$\bar{x}$	s
POST1	Basal	6.6818	2.7669
	DRTA	9.7727	2.7243
	Strat	7.7727	3.9271
POST2	Basal	5.5455	2.0407
	DRTA	6.2273	2.0915
	Strat	8.3636	2.9040
POST3	Basal	41.0455	5.6356
	DRTA	46.7273	7.3884
	Strat	44.2727	5.7668

```
Basal/POST1     DRTA/POST1     Strat/POST1
 1 |             1 |            1 | 0
 2 | 0           2 |            2 |
 3 | 0           3 |            3 | 0
 4 | 000         4 |            4 | 0000
 5 | 00000       5 | 00         5 | 00
 6 | 0           6 |            6 | 0
 7 | 00          7 | 000        7 | 000
 8 | 000         8 | 0000       8 | 0
 9 | 000         9 |            9 | 00
10 | 0          10 | 0000      10 |
11 |            11 | 00        11 | 00
12 | 00         12 | 000       12 | 00
               13 | 00        13 | 0
               14 | 00        14 | 0
                              15 | 0
```

```
Basal/POST2     DRTA/POST2        Strat/POST2
 0 |             0 | 0             0 |
 1 |             1 |               1 | 0
 2 |             2 |               2 |
 3 | 000         3 | 0             3 |
 4 | 00000       4 |               4 | 0
 5 | 000000      5 | 00            5 | 00
 6 | 0           6 | 0000000000    6 | 0
 7 | 00          7 | 0000          7 | 00
 8 | 000         8 | 00            8 | 000
 9 | 0           9 | 0             9 | 0000
10 | 0          10 |              10 | 000
               11 | 0            11 | 00
                                 12 | 00
                                 13 | 0
```

```
Basal/POST3     DRTA/POST3     Strat/POST3
3 |             3 | 01         3 |
3 | 223         3 |            3 | 33
3 | 5           3 |            3 | 4
3 | 66          3 | 7          3 |
3 | 99          3 |            3 | 8
4 | 0011        4 | 01         4 | 1
4 | 23          4 | 23         4 | 2223
4 | 4555        4 |            4 | 455
4 | 66          4 | 7          4 | 7
4 | 9           4 | 8889999    4 | 888999
5 |             5 | 0          5 | 01
5 |             5 | 33         5 | 3
5 | 4           5 | 455
                5 | 7
```

Minitab output: Analysis of Variance on POST1

```
Source    DF      SS      MS      F       p
Method     2   108.1    54.1   5.32   0.007
Error     63   640.5    10.2
Total     65   748.6
```

Analysis of Variance on POST2

```
Source    DF      SS      MS      F       p
Method     2   95.12   47.56   8.41   0.001
Error     63  356.41    5.66
Total     65  451.53
```

Analysis of Variance on POST3

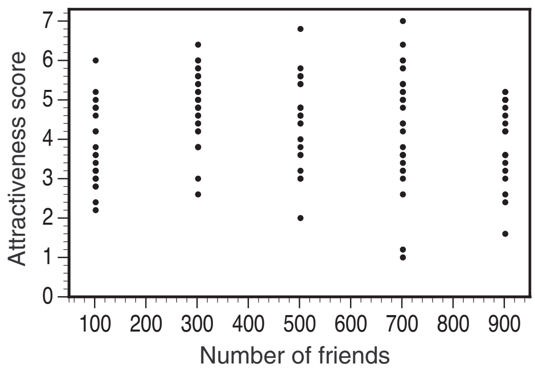
```
Source    DF      SS      MS      F       p
Method     2   357.3   178.7   4.48   0.015
Error     63  2511.7    39.9
Total     65  2869.0
```

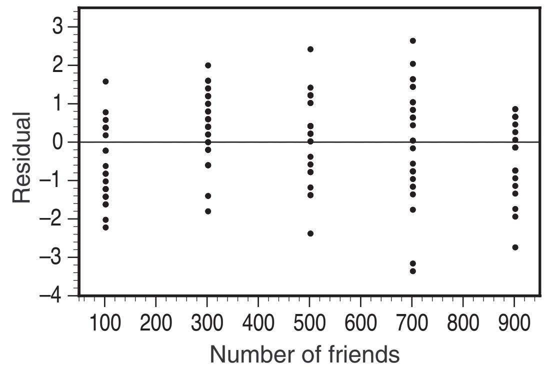
12.63. The scatterplot (below, left) suggests that a straight line is *not* the best choice of a model. Regression gives the formula

$$\widehat{\text{Score}} = 4.432 - 0.000102 \text{ Friends}$$

Not surprisingly, the slope is not significantly different from 0 ($t = -0.28$, $P = 0.782$). The regression only explains 0.1% of the variation in score. The residual plot (below, right) is nearly identical to the first scatterplot, and suggests (as that did) that a quadratic model might be a better choice.

Note: *If one fits a quadratic model, it does better (and has significant coefficients), but it still only explains 8.3% of the variation in attractiveness.*





Minitab output: Regression of attractiveness score on number of friends

```
The regression equation is Score = 4.43 -0.000102 Friends

Predictor        Coef       Stdev    t-ratio        p
Constant       4.4321      0.2060      21.51    0.000
Friends    -0.0001023   0.0003694      -0.28    0.782

s = 1.150        R-sq = 0.1%      R-sq(adj) = 0.0%
```

12.64. The pooled standard deviation s_p is found by looking at the spread of each observation about its *group* mean $\bar{x}_i$. The "total" standard deviation s given in Exercise 12.30 is the spread about the grand mean (the mean of all the data values, ignoring distinctions between groups). When we ignore group differences, we have more variation (uncertainty) in our data, so s is *almost always* larger than s_p.

This can be made clearer (to sufficiently mathematical students) by noting that the total variance s^2 can be found in the ANOVA table:

$$\text{Just as } s_p^2 = \frac{\text{SSE}}{\text{DFE}} = \text{MSE}, \; s^2 = \frac{\text{SST}}{\text{DFT}} = \text{MST}.$$

(The total mean square is not included in the ANOVA table but is easily computed from the values on the bottom line.) Because SSM + SSE = SST, we always have SSE $\leq$ SST, with equality only when the model is completely worthless (that is, when all group means equal the grand mean, so that SSM = 0). Because DFE < DFT, it might be that MSE $\geq$ MST but that does not happen very often.

12.66. With $\sigma = 7$ and means $\mu_1 = 40$, $\mu_2 = 47$, and $\mu_3 = 43$, we have $\bar{\mu} = \frac{40+47+43}{3} = 43.\overline{3}$ and noncentrality parameter:

$$\lambda = \frac{n\sum(\mu_i - \bar{\mu})^2}{\sigma^2} = \frac{(10)\left[(40-43.\overline{3})^2 + (47-43.\overline{3})^2 + (43-43.\overline{3})^2\right]}{49} = \frac{(10)(24.\overline{6})}{49} \doteq 5.0340$$

(The value of λ in the G•Power output below is slightly different due to rounding.) The degrees of freedom and critical value are the same as in Example 12.27: df 2 and 27, $F^* = 3.35$. Software reports the power as about 46%. Samples of size 10 are not adequate for this alternative; we should increase the sample size so that we have a better chance of detecting it. (For example, samples of size 20 give nearly 80% power for this alternative.)

G•Power output

```
Post-hoc analysis for "F-Test (ANOVA)", Global, Groups: 3:
Alpha: 0.0500
Power (1-beta): 0.4606
Effect size "f": 0.4096
Total sample size: 30
Critical value: F(2,27) = 3.3541
Lambda: 5.0332
```

12.67. **(a)** Sampling plans will vary but should attempt to address how cultural groups will be determined: Can we obtain such demographic information from the school administration? Do we simply select a large sample then poll each student to determine if he or she belongs to one of these groups? **(b)** Answers will vary with choice of H_a and desired power. For example, with the alternative $\mu_1 = \mu_2 = 4.4$, $\mu_3 = 5$, and standard deviation $\sigma = 1.2$, three samples of size 75 will produce power 0.89. (See G•Power output below.) **(c)** The report should make an attempt to explain the statistical issues involved; specifically, it should convey that sample sizes are sufficient to detect anticipated differences among the groups.

G•Power output

```
Post-hoc analysis for "F-Test (ANOVA)", Global, Groups: 3:
Alpha: 0.0500
Power (1-beta): 0.8920
Effect size "f": 0.2357
Total sample size: 225
Critical value: F(2,222) = 3.0365
Lambda: 12.4998
```

12.68. Recommended sample sizes will vary with choice of H_a and desired power. For example, with the alternative $\mu_1 = \mu_2 = 0.22$, $\mu_3 = 0.24$, and standard deviation $\sigma = 0.015$, three samples of size 10 will produce power 0.84, and samples of size 15 increase the power to 0.96. (See G•Power output below.) The report should make an attempt to explain the statistical issues involved; specifically, it should convey that sample sizes are sufficient to detect anticipated differences among the groups.

G•Power output

```
Post-hoc analysis for "F-Test (ANOVA)", Global, Groups: 3:
Alpha: 0.0500
Power (1-beta): 0.8379
Effect size "f": 0.6285
Total sample size: 30
Critical value: F(2,27) = 3.3541
Lambda: 11.8504
Note: Accuracy mode calculation.

Post-hoc analysis for "F-Test (ANOVA)", Global, Groups: 3:
Alpha: 0.0500
Power (1-beta): 0.9622
Effect size "f": 0.6285
Total sample size: 45
Critical value: F(2,42) = 3.2199
Lambda: 17.7756
```

12.69. The design can be similar, although the types of music might be different. Bear in mind that spending at a casual restaurant will likely be less than at the restaurants examined in Exercise 12.28; this might also mean that the standard deviations could be smaller. A pilot study might be necessary to get an idea of the size of the standard deviations. Decide how big a difference in mean spending you would want to detect, then do some power computations.

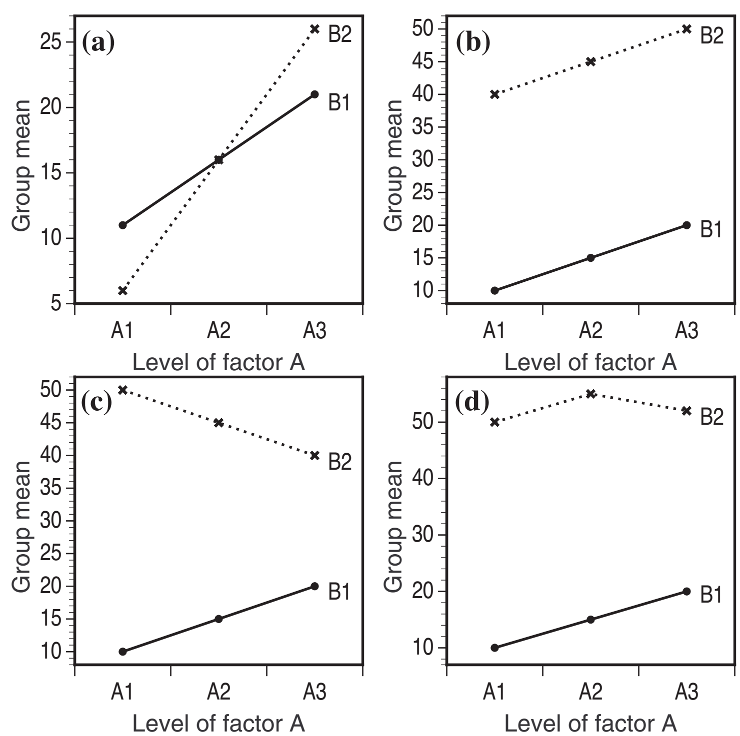
Chapter 13 Solutions

13.1. **(a)** Two-way ANOVA is used when there are two factors (explanatory variables). (The outcome [response] variable is assumed to have a Normal distribution, meaning that it can take any value, at least in theory.) **(b)** Each level of A should occur with all three levels of B. (Level A has two factors.) **(c)** The RESIDUAL part of the model represents the error. **(d)** DFAB $= (I-1)(J-1)$.

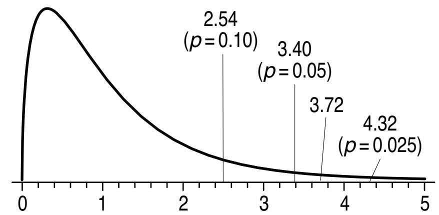
13.2. **(a)** Parallel profiles imply that there is *no* interaction. **(b)** It is not necessary that all sample sizes be the same. (The *standard deviations* must all be the same.) **(c)** s_p^2 is found by pooling the sample variances for each SRS. **(d)** The main effects can give useful information even in the presence of an interaction.

13.3. **(a)** A *large* value of the AB F statistic indicates that we should reject the hypothesis of no interaction. **(b)** The relationship is backwards: *Mean* squares equal *sum of* squares divided by degrees of freedom. **(c)** Under H_0, the ANOVA test statistics have an F distribution. **(d)** If the sample sizes are not the same, the sums of squares may not add for "some methods of analysis." (See the 'Caution' on page 680; for more detail, see `http://afni.nimh.nih.gov/sscc/gangc/SS.html`.)

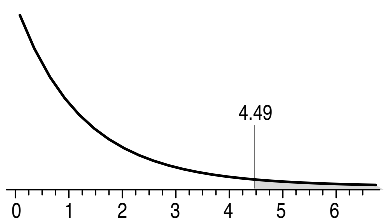
13.4. **(a)** Yes: The factor-A means change more drastically under B2 than under B1.
(b) No interaction (the lines are perfectly parallel).
(c) Yes: The factor-A means increase under B1, and decrease under B2.
(d) Yes: When A changes from level 2 to level 3, the means increase under B1 and decrease under B2.

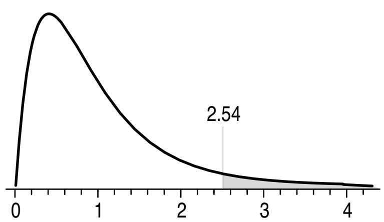


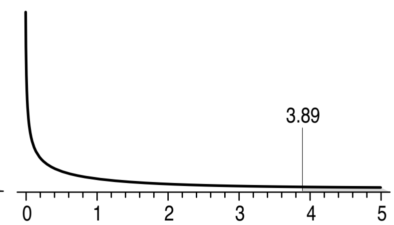
13.5. A 3×2 ANOVA with 5 observations per cell has $I = 3$, $J = 2$, and $N = 30$. **(a)** The degrees of freedom for interaction are $\text{DFAB} = (I-1)(J-1) = 2$ and $\text{DFE} = N - IJ = 24$. The five critical values from Table E are 2.54, 3.40, 4.32, 5.61, and 9.34. **(b)** The sketch on the right shows the observed F-value given in part (c) and the bounding critical values from Table E. **(c)** In Table E, we see that $3.40 < F < 4.32$, so $0.025 < P < 0.05$. (Software gives 0.0392.) **(d)** The mean profiles would not look parallel because the interaction term is significantly different from 0.



13.6. The answers are found in Table E (or using software) with $P = 0.05$. **(a)** We have $I = 2$, $J = 4$ and $N = 24$, so $\text{DFA} = 1$ and $\text{DFE} = 16$. We would reject H_0 if $F > 4.49$ (software gives 4.4940). **(b)** We have $I = J = 4$ and $N = 32$, so $\text{DFAB} = 9$ and $\text{DFE} = 16$. We would reject H_0 if $F > 2.54$ (software: 2.5377). **(c)** We have $I = J = 2$ and $N = 204$, so $\text{DFAB} = 1$ and $\text{DFE} = 200$. We would reject H_0 if $F > 3.89$ (software: 3.8884).





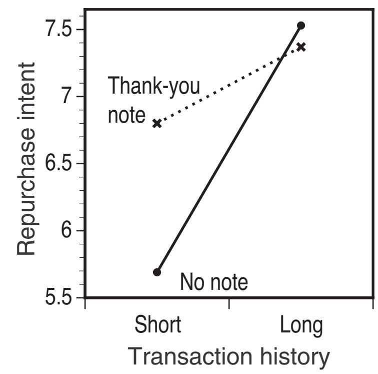


13.7. **(a)** The factors are gender ($I = 2$) and age ($J = 3$). The response variable is the percent of pretend play. The total number of observations is $N = (2)(3)(11) = 66$. **(b)** The factors are time after harvest ($I = 5$) and amount of water ($J = 2$). The response variable is the percent of seeds germinating. The total number of observations is $N = 30$ (3 lots of seeds in each of the 10 treatment combinations). **(c)** The factors are mixture ($I = 6$) and freezing/thawing cycles ($J = 3$). The response variable is the strength of the specimen. The total number of observations is $N = 54$. **(d)** The factors are training programs ($I = 4$) and the number of days to give the training ($J = 2$). The response variable is not specified, but presumably is some measure of the training's effectiveness. The total sample size is $N = 80$.

13.8. The table on the right summarizes the degrees of freedom for each source.

		(a)	**(b)**	**(c)**	**(d)**
	$I =$	2	5	6	4
	$J =$	3	2	3	2
Source	$N =$	66	30	54	80
A	$I - 1 =$	1	4	5	3
B	$J - 1 =$	2	1	2	1
AB	$(I-1)(J-1) =$	2	4	10	3
Error	$N - IJ =$	60	20	36	72

13.9. **(a)** There appears to be an interaction: A thank-you note increases repurchase intent by over 1 point for those with short history, and decreases it (very slightly) for customers with long history. Note that either variable could be on the horizontal axis in the plot of means. **(b)** The marginal means are



Short history	6.245	No thank-you note	6.61
Long history	7.45	Thank-you note	7.085

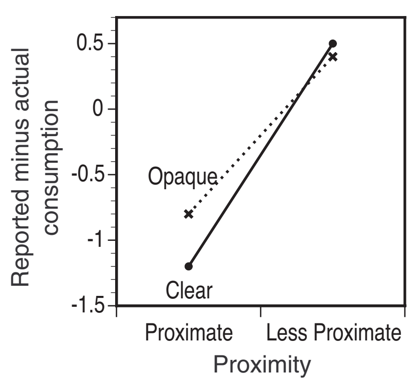
For example, $\frac{5.69+6.80}{2} = 6.245$. The history marginal means convey the fact that repurchase intent is higher for customers with long history. The thank-you note marginal means suggest that a thank-you note increases repurchase intent, but they are harder to interpret because of the interaction.

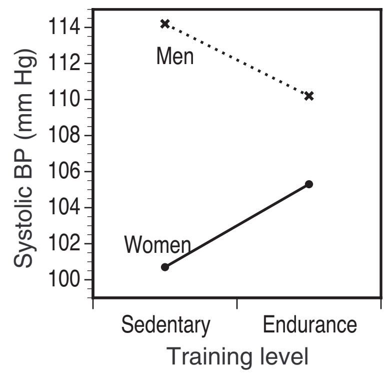
13.10. With $I = J = 2$ levels for each factor, the three missing entries in the DF column are all 1. The MS entries are computed as $\frac{\text{MS}}{\text{DF}}$, and the F statistics are $\frac{\text{MS}}{\text{DFE}}$. Comparing each test statistic to an $F(1, 160)$ distribution gives the P-values.

Source	DF	SS	MS	F	P-value
Transaction history	1	61.445	61.445	12.94	0.0004
Thank-you statement	1	21.810	21.810	4.59	0.0336
Interaction	1	15.404	15.404	3.24	0.0736
Error	160	759.904	4.7494		

The interaction is not quite significant, but the two main effects are.

13.11. **(a)** The plot suggests a possible interaction because the means are not parallel. (Note that we could have chosen to put dish type on the horizontal axis instead of proximity; either explanatory variable will do.) **(b)** By subjecting the same individual to all four treatments, rather than four individuals to one treatment each, we reduce the within-groups variability (the residual), which makes it easier to detect between-groups variability (the main effects and interactions).



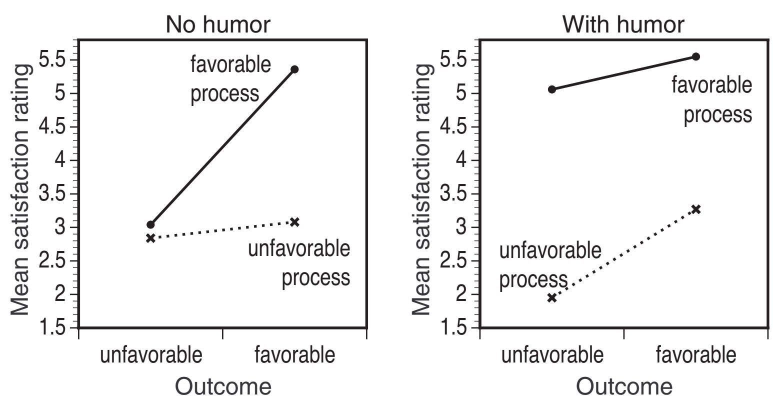


13.12. (a) The plot suggests a gender effect: Men had higher postexercise blood pressure (BP) than women. There also appears to be an interaction: BP was higher for endurance-trained women than for sedentary women (as the researchers had hypothesized), but for men, that pattern was reversed. **(b)** The complete ANOVA table is given in the Minitab output below. The apparent interaction noted in (a) was not significant, but there is a significant effect of gender. **(c)** Subjects with a high before-exercise BP are likely to have higher postexercise BP, as well. By incorporating both measurements, the researchers can focus of the *change* in BP after exercising, which should be a better measure of the effect of exercise.

Note: *The fact that the interaction was not significant, despite the appearance of the plot, is due to the large variation in individual BPs, indicated by the sizes of the standard errors given in the table. (Observe that these are standard* errors, *not standard* deviations.*) One reason for measuring change in BP—as suggested in (c)—is that we might expect this measurement to have less subject-to-subject variation.*

Minitab output: Two-way ANOVA for BP on gender and training level

Source	DF	SS	MS	F	P
Gender	1	677.12	677.12	7.65	0.010
Training	1	0.72	0.72	0.01	0.929
Gender*Training	1	147.92	147.92	1.67	0.207
Error	28	2478.00	88.50		
Total	31	3303.76			



13.13. (a) There may be an interaction: For a favorable process, a favorable outcome increases satisfaction quite a bit more than for an unfavorable process (+2.32 versus +0.24). **(b)** With humor, the increase in satisfaction from a favorable outcome is less for a favorable process (+0.49 compared to +1.32). **(c)** There seems to be a three-factor interaction, because the interactions in parts (a) and (b) are different.

13.14. For the pooled standard deviation, we first find

$$s_p^2 = \frac{(26)(0.79^2) + (28)(0.47^2) + \cdots + (29)(0.71^2)}{26 + 28 + \cdots + 29} = \frac{88.6838}{233} \doteq 0.3806$$

so $s_p \doteq \sqrt{0.3806} \doteq 0.6169$. There were $N = 241$ students in the sample, and 8 groups, so this has df $= 241 - 8 = 233$. The largest-to-smallest standard deviation ratio is $\frac{0.79}{0.47} \doteq 1.68 < 2$, so it is reasonable to use this pooled estimate.

13.15. Marginal means are listed in the table below. In each case, we find the average of the four means for each level of the characteristic. For example, for Humor, we have

$$\text{No humor: } \frac{3.04 + 5.36 + 2.84 + 3.08}{4} = 3.58$$

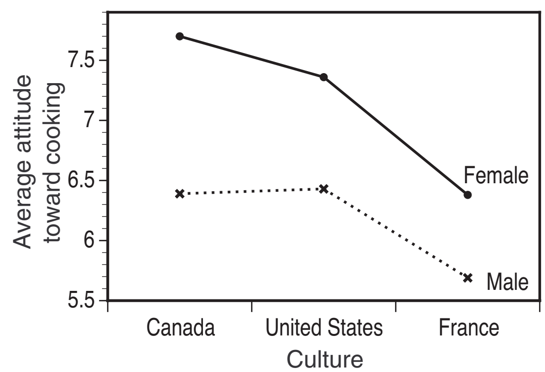
$$\text{Humor: } \frac{5.06 + 5.55 + 1.95 + 3.27}{4} = 3.9575$$

The presence of humor slightly increases mean satisfaction. The process and outcome effects appear to be greater (that is, the change in mean satisfaction is greater).

Marginal means

Humor		Process		Outcome	
No	3.58	Favorable	4.7525	Favorable	4.315
Yes	3.9575	Unfavorable	2.785	Unfavorable	3.2225

13.16. (a) Within a given culture, females generally have a more positive attitude toward cooking than males. Attitudes in France are less positive than those in the U.S. and Canada. **(b)** While the means plot is not perfectly parallel, it is not clear that it indicates an interaction. If an interaction *is* present, it is that the female/male difference is greatest in Canada, and least in France.



13.17. For the pooled standard deviation, we first find

$$s_p^2 = \frac{(237)(1.668^2) + (124)(1.909^2) + \cdots + (86)(1.875^2)}{237 + 124 + \cdots + 86} \doteq \frac{2535.19}{805} \doteq 3.1493$$

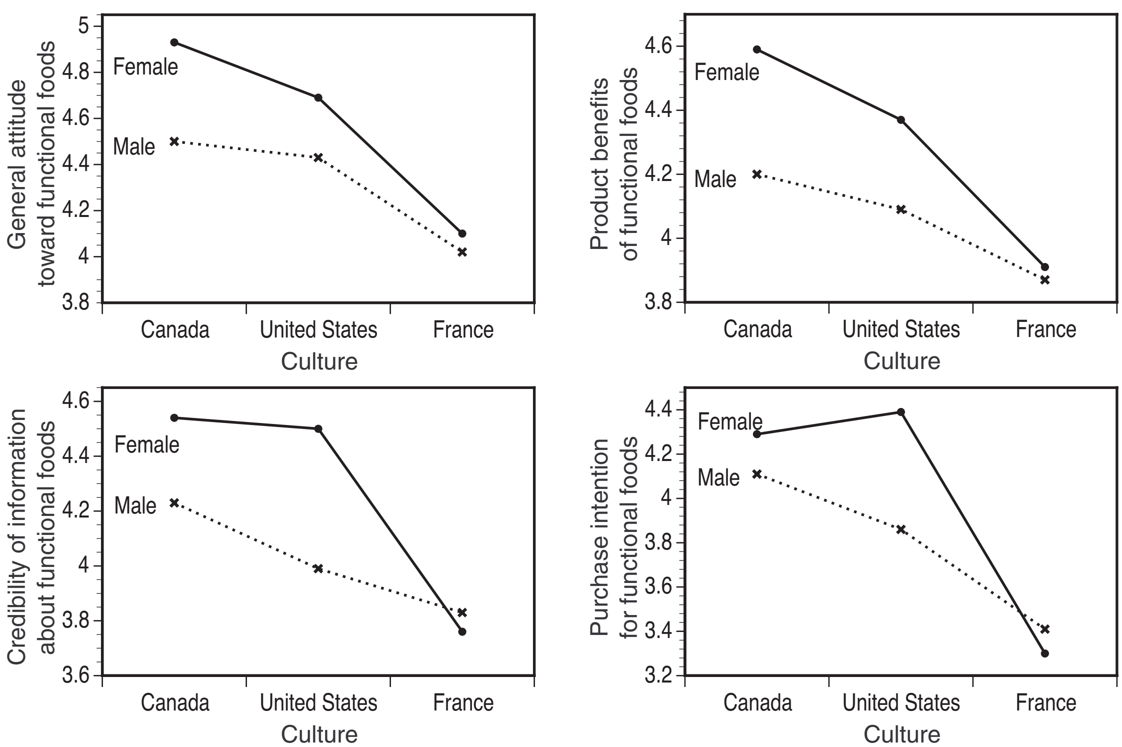
so $s_p \doteq \sqrt{3.1493} \doteq 1.7746$. The largest-to-smallest standard deviation ratio is $\frac{2.024}{1.601} \doteq 1.26 < 2$, so it is reasonable to use this pooled estimate.

13.18. (a) With a total sample size of $N = 811$, and six groups, we have df $= 805$. **(b)** There are 6 means, so there are $\frac{6 \cdot 5}{2} = 15$ comparisons. **(c)** Using $s_p \doteq 1.7746$ from the previous exercise, the t statistics are $t_{ij} = (\bar{x}_i - \bar{x}_j)/\text{SE}_{ij}$, where $\text{SE}_{ij} = s_p\sqrt{\frac{1}{n_i} + \frac{1}{n_j}}$. The complete set of differences, standard errors, and t values is listed on the following page; the eight largest differences (marked with an asterisk) are significant.

Note: *If doing these computations by hand, it is best to start with the largest differences and work down until one finds a difference that is not significant. (One should then check a few more, as there might be one or two more significant differences remaining because the standard errors vary with sample size.) In this set of data, for example, there is little reason to check for a significant difference between the Canadian male, U.S. male, and French female means.*

Means	Difference	SE	t	
Female/Canada – Male/Canada	1.31	0.1960	6.6828	*
Female/Canada – Female/U.S.	0.34	0.1759	1.9334	
Female/Canada – Male/U.S.	1.27	0.2107	6.0262	*
Female/Canada – Female/France	1.32	0.2272	5.8088	*
Female/Canada – Male/France	2.01	0.2223	9.0406	*
Male/Canada – Female/U.S.	−0.97	0.2071	−4.6839	*
Male/Canada – Male/U.S.	−0.04	0.2374	−0.1685	
Male/Canada – Female/France	0.01	0.2522	0.0397	
Male/Canada – Male/France	0.70	0.2478	2.8251	
Female/U.S. – Male/U.S.	0.93	0.2211	4.2067	*
Female/U.S. – Female/France	0.98	0.2369	4.1376	*
Female/U.S. – Male/France	1.67	0.2321	7.1937	*
Male/U.S. – Female/France	0.05	0.2638	0.1895	
Male/U.S. – Male/France	0.74	0.2596	2.8508	
Female/France – M/France	0.69	0.2731	2.5262	

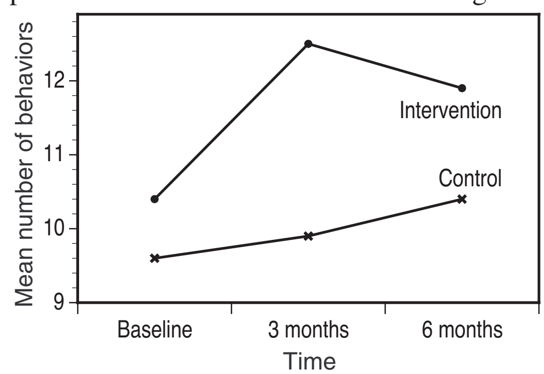
13.19. Means plots are below. Possible observations: Except for female responses to purchase intention, means decreased from Canada to the United States to France. Females had higher means than men in almost every case, except for French responses to credibility and purchase intention (suggesting a modest interaction). Gender differences in France are considerably smaller than in either Canada or the United States.



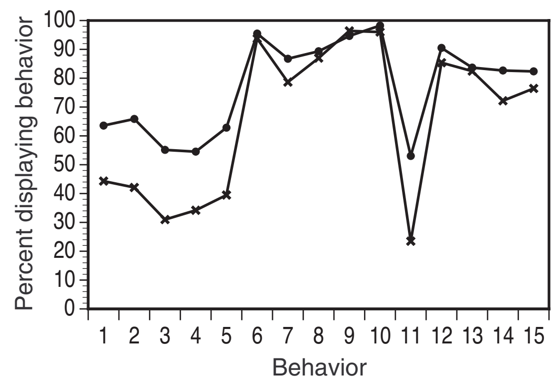
13.20. Opinions of undergraduate students might be similar to a large segment of the young adult population, but this sample is probably not an unbiased representation of that group. Filling out the surveys in class might also affect the usefulness of the responses in some way (although it is difficult to predict what that effect might be).

13.21. **(a)** The marginal means (as well as the individual cell means) are in the table below. The first two means suggest that the intervention group showed more improvement than the control group. **(b)** Interaction means that the mean number of actions changes differently over time for the two groups. We see this in the plot below because the lines connecting the means are not parallel.

	Time			
Group	Baseline	3 mo.	6 mo.	Mean
Intervention	10.4	12.5	11.9	11.6
Control	9.6	9.9	10.4	9.967
Mean	10.0	11.2	11.15	10.783



13.22. **(a)** The data might be displayed in a variety of ways. Because there are so many numbers (intervention and control groups, at baseline, 3 months, and 6 months), the graph can very easily become overwhelmingly crowded; in order to avoid this, the graph on the right shows the percentage for each group averaged over the three times. Any reasonable graphical display will likely be judged more effective than Table 13.1; for the most part, it is easier to interpret pictures than lists of numbers. **(b)** The behaviors seemed to fall into two categories: Those that both groups did most of the time and those that were less common. The biggest differences between the control and intervention groups are in the latter group, which includes the first five and the 11th behaviors: hide money, hide extra keys, abuse code to alert family, hide extra clothing, asked neighbors to call police, removed weapons. These behaviors should receive special attention in future programs. **(c)** Perhaps the results of this study may be less applicable to smaller communities, or to those which are less diverse.



13.23. We have $I = 3$, $J = 2$, and $N = 30$, so the degrees of freedom are DFA = 2, DFB = 1, DFAB = 2, and DFE = 24. This allows us to determine P-values (or to compare to Table E), and we find that there are no significant effects (although B is close):

$$F_A = 1.87 \text{ has df 2 and 24, so } P = 0.1759$$

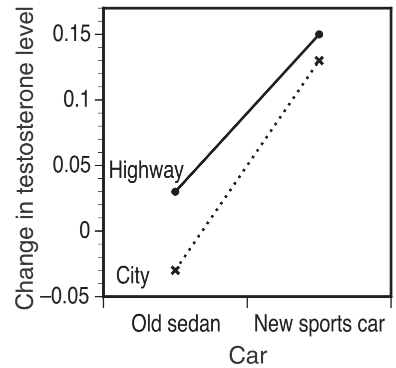
$$F_B = 3.49 \text{ has df 1 and 24, so } P = 0.0740$$

$$F_{AB} = 2.14 \text{ has df 2 and 24, so } P = 0.1396$$

13.24. **(a)** Based on the given P-values, the interaction and the main effect of B are significant at $\alpha = 0.05$. **(b)** In order to summarize the results, we would need to know the number of levels for each factor (I and J) and the sample sizes in each cell (n_{ij}). We also would want to know the sample cell means $\bar{x}_{ij}$ so that we could interpret the significant main effect and the nature of the interaction.

13.25. **(a)** The means are nearly parallel, and show little evidence of an interaction. **(b)** With equal sample sizes, the pooled variance is simply the unweighted average of the variances:

$$s_p^2 = \frac{1}{4}(0.12^4 + 0.14^2 + 0.12^2 + 0.13^2) = 0.016325$$



Therefore, $sp = \sqrt{0.016325} \doteq 0.1278$. **(c)** Note that all of these contrasts have been arranged so that, if the researchers' suspicions are correct, the contrast will be positive. To compare new-car testosterone change to old-car change, the appropriate contrast is

$$\psi_1 = \tfrac{1}{2}(\mu_{\text{new,city}} + \mu_{\text{new,highway}}) - \tfrac{1}{2}(\mu_{\text{old,city}} + \mu_{\text{old,highway}})$$

To compare city change to highway change for new cars, we take

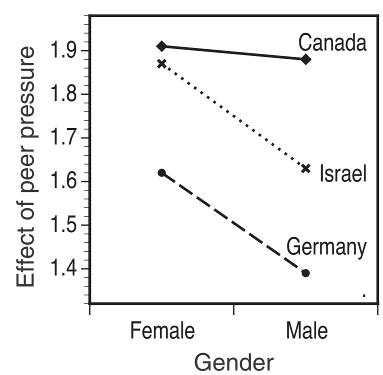
$$\psi_2 = \mu_{\text{new,city}} - \mu_{\text{new,highway}}$$

To compare highway change to city change for old cars, we take

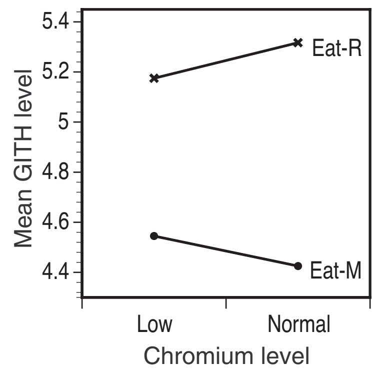
$$\psi_3 = \mu_{\text{old,highway}} - \mu_{\text{old,city}}$$

(d) By subjecting the same individual to all four treatments, rather than four individuals to one treatment each, we reduce the within-groups variability (the residual), which makes it easier to detect between-groups variability (the main effects and interactions).

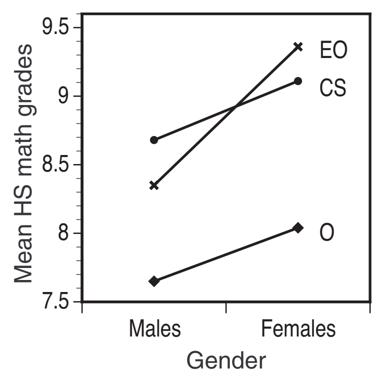
13.26. **(a)** The significant interaction ($P = 0.016$) is visible in the plot of means: males and females in Canada report similar amounts of peer pressure, while males in Germany and Israel report less peer pressure than females. (Note that we could have chosen to put country on the horizontal axis instead of gender.) **(b)** It is not entirely clear how the value placed on achievement relates to fear of being called a nerd or teacher's pet. One view might be that in a culture that values achievement, peers should be less likely to "punish" those who do well. If we take this view, the data would refute the researchers' hypothesis. One could also argue that placing a high value on achievement might make students more competitive, and name-calling might be an expression of that competitiveness. Under this view, the low peer-pressure scores for Germany support the researchers' hypothesis. **(c)** With both responses, we could explore the relationship between achievement and fear of peer pressure—for example, are high-achieving students more or less concerned about peer pressure than low-achieving students?



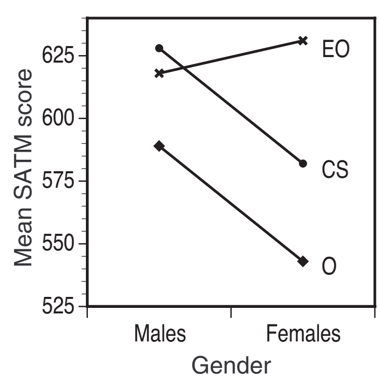
13.27. **(a)** Plot on the right. **(b)** There seems to be a fairly large difference between the means based on how much the rats were allowed to eat, but not very much difference based on the chromium level. There may be an interaction: the NM mean is lower than the LM mean, while the NR mean is higher than the LR mean. **(c)** The marginal means are L: 4.86, N: 4.871, M: 4.485, R: 5.246. For low chromium level (L), R minus M is 0.63; for normal chromium (N), R minus M is 0.892. Mean GITH levels are lower for M than for R; there is not much difference for L versus N. The difference between M and R is greater among rats who had normal chromium levels in their diets (N).



13.28. The "Other" category had the lowest mean HS math grades for both genders; this is apparent from the graph (right) and from the marginal means (CS: 8.895, EO: 8.855, O: 7.845). Females had higher mean grades; the female marginal mean is $8.83\overline{6}$ compared to $8.22\overline{6}$ for males. The female − male difference is similar for CS and O (about 0.5) but is about twice as big for EO (an interaction).



13.29. The "Other" category had the lowest mean SATM score for both genders; this is apparent from the graph (right) and from the marginal means (CS: 605, EO: 624.5, O: 566.) Males had higher mean scores in CS and O, while females are slightly higher in EO; this indicates an interaction. Overall, the marginal means by gender are 611.7 (males) and 585.3 (females).

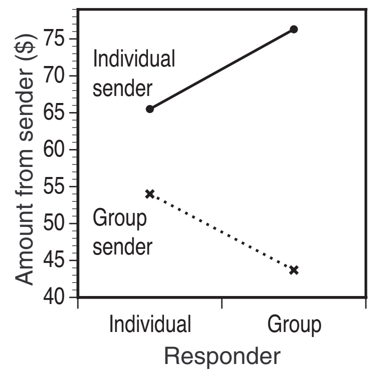


13.30. A study today might include a category for those who declared a major such as Information Technology (which probably did not exist at the time of the initial study). Some variables that might be useful to consider: grade in first programming course, high school physics grades, etc.

13.31. **(a)** The pooled variance is

$$s_p^2 = \frac{(31)(36.4^2) + (24)(31.2^2) + (24)(41.6^2) + (26)(42.4^2)}{31 + 24 + 24 + 26}$$
$$= \frac{152,711.52}{105} \doteq 1454.4$$

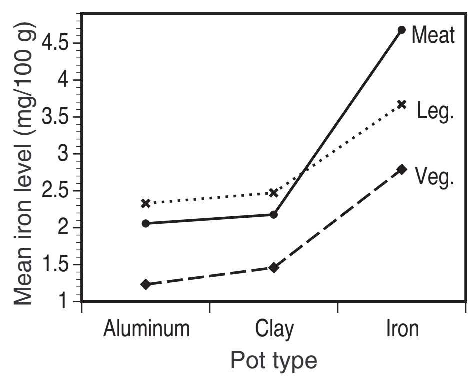
so $s_p \doteq \sqrt{1454.4} \doteq 38.14$. There were $N = 109$ items in the sample, and four groups, so df $= 105$. **(b)** Pooling is reasonable because the ratio of the largest and smallest standard deviations is $\frac{42.4}{31.2} \doteq 1.36 < 2$. **(c)** The marginal means are:



Sender		Responder	
Individual:	$\frac{1}{2}(65.5 + 76.3) = \$70.90$	Individual:	$\frac{1}{2}(65.5 + 54.0) = \$59.75$
Group:	$\frac{1}{2}(54.0 + 43.7) = \$48.85$	Group:	$\frac{1}{2}(76.3 + 43.7) = \$60.00$

(d) There appears to be an interaction: Individuals send more money to groups, while groups send more money to individuals. **(e)** Compare the statistics to an $F(1, 105)$ distribution. The three P-values are 0.0033 (sender), 0.9748 (responder), and 0.1522 (interaction). Only the main effect of sender is significant.

13.32. **(a)** The sample size is $n = 4$ for each pot/food combination; means and standard deviations are given in the table below. The largest-to-smallest ratio is $0.63/0.07 \doteq 8.8$, which is well above our guideline for pooling. **(b)** The iron levels differed among the three food types, and for all food types, aluminum and clay pots produced similar iron levels, while iron pots resulted in much higher iron levels. There is also evidence of an interaction: Iron levels in iron pots rose much more for meat than for legumes or vegetables. **(c)** The ANOVA table (below) shows that all three effects are quite significant.



Iron (mg per 100 g)

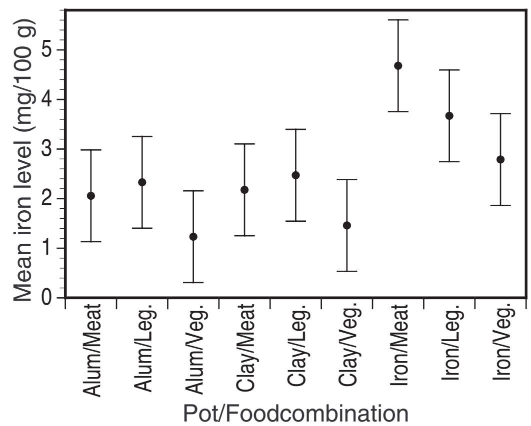
	Meat		Legumes		Vegetables	
Pot type	$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	s
Aluminum	2.0575	0.2520	2.3300	0.1111	1.2325	0.2313
Clay	2.1775	0.6213	2.4725	0.0714	1.4600	0.4601
Iron	4.6800	0.6283	3.6700	0.1726	2.7900	0.2399

Minitab output: Two-way ANOVA for iron on pot and food

```
Source     DF        SS        MS      F      P
Pot         2   24.8940   12.4470  92.26  0.000
Food        2    9.2969    4.6484  34.46  0.000
Pot*Food    4    2.6404    0.6601   4.89  0.004
Error      27    3.6425    0.1349
Total      35   40.4738
```

13.33. Yes; the iron-pot means are the highest, and the F statistic for testing the effect of the pot type is very large. (In this case, the interaction does not weaken any evidence that iron-pot foods contain more iron; it only suggests that while iron pots increase iron levels in all foods, the effect is strongest for meats.)

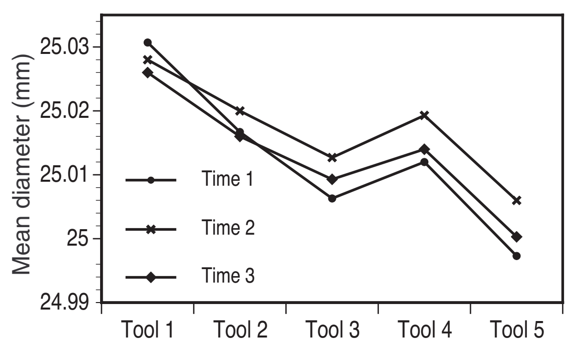
13.34. The ANOVA table (below) shows significant evidence that at least one group mean is different. With df $= 27$, 36 comparisons, and $\alpha = 0.05$, the Bonferroni critical value is $t^{**} = 3.5629$. The pooled standard deviation is $s_p \doteq 0.3673$, the standard error of each difference $\mathrm{SE}_D = s_p\sqrt{1/4 + 1/4} \doteq 0.2597$, so two means are significantly different if they differ by $t^{**}\mathrm{SE}_D \doteq 0.9253$. The "error bars" in the plot on the right are drawn with this length (above and below each mean), so two means are significantly different if the "dot" for one mean does not fall within the other mean's error bars. For example, we find that iron/meat is significantly larger than everything else, and iron/legumes is significantly different from everything except iron/vegetable. These conclusions are consistent with the results of the two-way ANOVA.



Minitab output: One-way ANOVA for iron on pot/food combinations

Source	DF	SS	MS	F	p
Potfood	8	36.831	4.604	34.13	0.000
Error	27	3.643	0.135		
Total	35	40.474			

13.35. **(a)** For all tool/time combinations, $n = 3$. Means and standard deviations are in the table on the following page. Note that five cells had no variability ($s = 0$). **(b)** Plot on the right. Except for tool 1, mean diameter is highest at time 2. Tool 1 had the highest mean diameters, followed by tool 2, tool 4, tool 3, and tool 5. **(c)** Minitab output below; all F statistics are highly significant. **(d)** There is strong evidence of a difference in mean diameter among the tools and among the times. There is also an interaction (specifically, tool 1's mean diameters changed differently over time compared to the other tools).



Minitab output: Two-way ANOVA for diameter on tool and time

Source	DF	SS	MS	F	P
Tool	4	0.00359720	0.00089930	412.94	0.000
Time	2	0.00018991	0.00009496	43.60	0.000
Tool*Time	8	0.00013320	0.00001665	7.65	0.000
Error	30	0.00006533	0.00000218		
Total	44	0.00398564			

Diameter (mm)

	Time 1 (8:00AM)		Time 2 (11:00AM)		Time 3 (3:00PM)	
Tool	$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	s
1	25.0307	0.001155	25.0280	0	25.0260	0
2	25.0167	0.001155	25.0200	0.002000	25.0160	0
3	25.0063	0.001528	25.0127	0.001155	25.0093	0.001155
4	25.0120	0	25.0193	0.001155	25.0140	0.004000
5	24.9973	0.001155	25.0060	0	25.0003	0.001528

13.36. All means and standard deviations will change by a factor of 0.04; the plot is identical to that in Exercise 13.35, except that the vertical scale is different. All SS and MS values change by a factor of $0.04^2 = 0.0016$, but the F- (and P-) values are the same.

Minitab output: Two-way ANOVA for diameter on tool and time

```
Source       DF          SS          MS        F      P
Tool          4   0.0000058   0.0000014   412.94  0.000
Time          2   0.0000003   0.0000002    43.60  0.000
Tool*Time     8   0.0000002   0.0000000     7.65  0.000
Error        30   0.0000001   0.0000000
Total        44   0.0000064
```

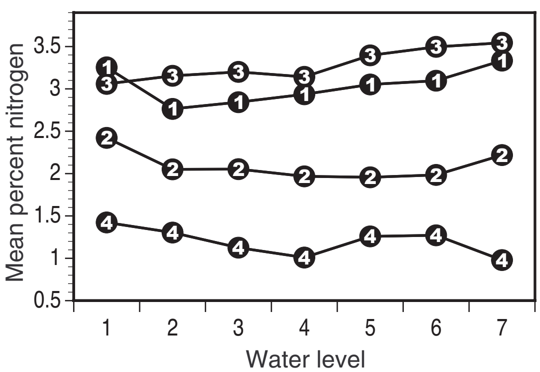
13.37. (a) All three F values have df 1 and 945, and the P-values are < 0.001, < 0.001, and 0.1477. Gender and handedness both have significant effects on mean lifetime, but there is no significant interaction. **(b)** Women live about 6 years longer than men (on the average), while right-handed people average 9 more years of life than left-handed people. "There is no interaction" means that handedness affects both genders in the same way, and vice versa.

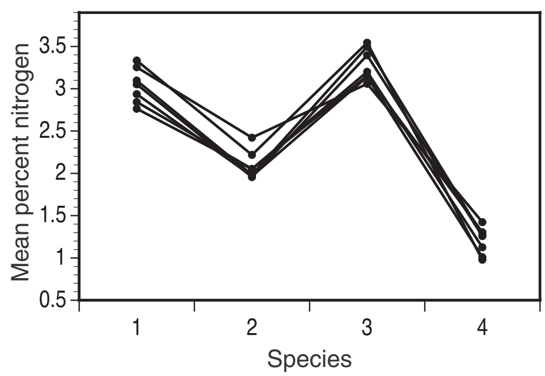
13.38. (a) $F_{\text{series}} = 7.02$ with df 3 and 61; this has $P = 0.0004$. $F_{\text{holder}} = 1.96$ with df 1 and 61; this has $P = 0.1665$. $F_{\text{interaction}} = 1.24$ with df 3 and 61; this has $P = 0.3026$. Only the series had a significant effect; the presence or absence of a holder and series/holder interaction did not significantly affect the mean radon reading. **(b)** Because the ANOVA indicates that these means are significantly different, we conclude that detectors produced in different production runs give different readings for the same radon level. This inconsistency may indicate poor quality control in production.

Note: *In the initial printing of the text, the total sample size ($N = 69$) was not given, without which we cannot determine the denominator degrees of freedom for part (a).*

13.39. (a) & **(b)** The table below lists the means and standard deviations (the latter in parentheses) of the nitrogen contents of the plants. The two plots below suggest that plant 1 and plant 3 have the highest nitrogen content, plant 2 is in the middle, and plant 4 is the lowest. (In the second plot, the points are so crowded together that no attempt was made to differentiate among the different water levels.) There is no consistent effect of water level on nitrogen content. Standard deviations range from 0.0666 to 0.3437, for a ratio of 5.16—larger than we like. **(c)** Minitab output below. Both main effects and the interaction are highly significant.

	Amount of water per day						
Species	50mm	150mm	250mm	350mm	450mm	550mm	650mm
1	3.2543 (0.2287)	2.7636 (0.0666)	2.8429 (0.2333)	2.9362 (0.0709)	3.0519 (0.0909)	3.0963 (0.0815)	3.3334 (0.2482)
2	2.4216 (0.1654)	2.0502 (0.1454)	2.0524 (0.1481)	1.9673 (0.2203)	1.9560 (0.1571)	1.9839 (0.2895)	2.2184 (0.1238)
3	3.0589 (0.1525)	3.1541 (0.3324)	3.2003 (0.2341)	3.1419 (0.2965)	3.3956 (0.2533)	3.4961 (0.3437)	3.5437 (0.3116)
4	1.4230 (0.1738)	1.3037 (0.2661)	1.1253 (0.1230)	1.0087 (0.1310)	1.2584 (0.2489)	1.2712 (0.0795)	0.9788 (0.2090)

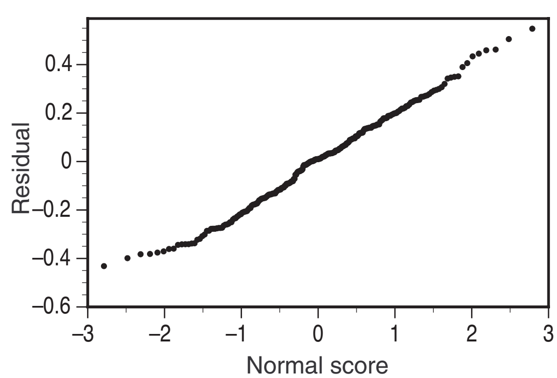




Minitab output: Two-way ANOVA for Pctnit on species and water

```
Source          DF        SS        MS       F      P
Species          3  172.3916   57.4639 1301.32  0.000
Water            6    2.5866    0.4311    9.76  0.000
Species*Water   18    4.7446    0.2636    5.97  0.000
Error          224    9.8914    0.0442
Total          251  189.6143
```

13.40. The residuals appear to be reasonably Normal, with no apparent outliers and no clear patterns.



13.41. For each water level, there is highly significant evidence of variation in nitrogen level among plant species (Minitab output below). For each water level, we have df $= 32$, 6 comparisons, and $\alpha = 0.05$, so the Bonferroni critical value is $t^{**} = 2.8123$. (If we take into account that there are 7 water levels, so that overall we are performing $6 \times 7 = 42$ comparisons, we should take $t^{**} = 3.5579$.) The table on the right gives the pooled standard deviations s_p, the standard errors of each difference $\text{SE}_D = s_p\sqrt{1/9 + 1/9}$, and the "minimum significant difference" MSD $= t^{**}\text{SE}_D$ (two means are significantly different if they differ by at least this amount). MSD1 uses $t^{**} = 2.8123$, and MSD2 uses $t^{**} = 3.5579$. As it happens, for either choice of MSD, the only *non*significant differences are between species 1 and 3 for water levels 1, 4, and 7. (These are the three closest pairs of points in the plot from the solution to Exercise 13.39.) Therefore, for every water level, species 4 has the lowest nitrogen level and species 2 is next. For water levels 1, 4, and 7, species 1 and 3 are statistically tied for the highest level; for the other levels, species 3 is the highest, with species 1 coming in second.

Water level	s_p	SE_D	MSD1	MSD2
1	0.1824	0.0860	0.2418	0.3059
2	0.2274	0.1072	0.3015	0.3814
3	0.1912	0.0902	0.2535	0.3208
4	0.1991	0.0939	0.2640	0.3340
5	0.1994	0.0940	0.2643	0.3344
6	0.2318	0.1093	0.3073	0.3887
7	0.2333	0.1100	0.3093	0.3913

Minitab output: One-way ANOVA on species for water level 1

```
Source    DF        SS        MS        F        p
Species    3   18.3711    6.1237   184.05    0.000
Error     32    1.0647    0.0333
Total     35   19.4358
```

One-way ANOVA on species for water level 2

```
Source    DF        SS        MS        F        p
Species    3   17.9836    5.9945   115.93    0.000
Error     32    1.6546    0.0517
Total     35   19.6382
```

One-way ANOVA on species for water level 3

```
Source    DF        SS        MS        F        p
Species    3   22.9171    7.6390   208.87    0.000
Error     32    1.1704    0.0366
Total     35   24.0875
```

One-way ANOVA on species for water level 4

```
Source    DF        SS        MS        F        p
Species    3   25.9780    8.6593   218.37    0.000
Error     32    1.2689    0.0397
Total     35   27.2469
```

One-way ANOVA on species for water level 5

```
Source    DF        SS        MS        F        p
Species    3   26.2388    8.7463   220.01    0.000
Error     32    1.2721    0.0398
Total     35   27.5109
```

One-way ANOVA on species for water level 6

```
Source    DF        SS        MS        F        p
Species    3   28.0648    9.3549   174.14    0.000
Error     32    1.7191    0.0537
Total     35   29.7838
```

One-way ANOVA on species for water level 7

```
Source    DF        SS        MS        F        p
Species    3   37.5829   12.5276   230.17    0.000
Error     32    1.7417    0.0544
Total     35   39.3246
```

13.42. The F statistics for all four ANOVAs are significant, and all four regressions are significant as well. However, the regressions all have low R^2 (varying from 6.4% to 27.3%), and plots indicate that a straight line is not really appropriate except perhaps for plant 3 (which had the highest R^2 value).

Minitab output: One-way ANOVA on water for plant species 1

```
Source     DF        SS        MS        F        p
Water       6    2.3527    0.3921    14.25    0.000
Error      56    1.5413    0.0275
Total      62    3.8940
```

One-way ANOVA on water for plant species 2

```
Source     DF        SS        MS        F        p
Water       6    1.5626    0.2604     7.51    0.000
Error      56    1.9420    0.0347
Total      62    3.5046
```

One-way ANOVA on water for plant species 3

```
Source     DF        SS        MS        F        p
Water       6    1.9764    0.3294     4.15    0.002
Error      56    4.4464    0.0794
Total      62    6.4228
```

One-way ANOVA on water for plant species 4

```
Source     DF        SS        MS        F        p
Water       6    1.4396    0.2399     6.85    0.000
Error      56    1.9618    0.0350
Total      62    3.4013
```

Regression of plant species 1 on water

```
The regression equation is plant1 = 2.88 + 0.0397 Water

Predictor       Coef       Stdev    t-ratio        p
Constant     2.88097     0.06745      42.71    0.000
Water        0.03971     0.01508       2.63    0.011

s = 0.2394      R-sq = 10.2%     R-sq(adj) = 8.7%
```

Regression of plant species 2 on water

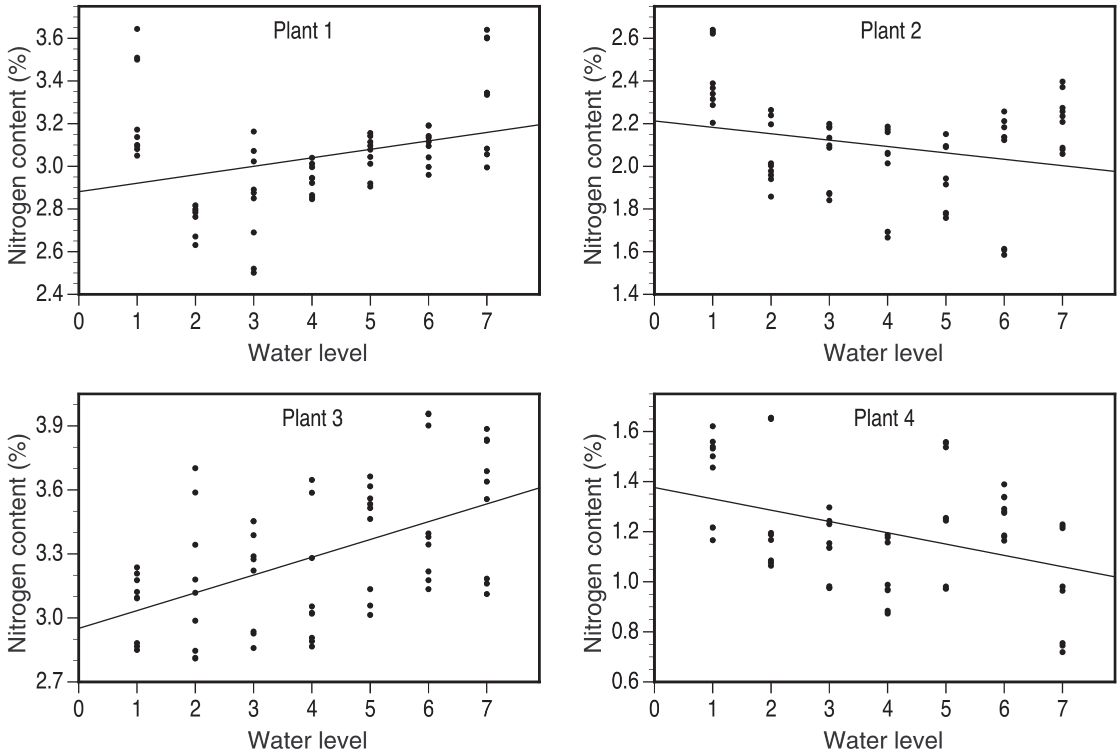
```
The regression equation is plant2 = 2.21 - 0.0299 Water

Predictor       Coef       Stdev    t-ratio        p
Constant     2.21262     0.06531      33.88    0.000
Water       -0.02994     0.01460      -2.05    0.045

s = 0.2318      R-sq = 6.4%      R-sq(adj) = 4.9%
```

Regression of plant species 3 on water

```
The regression equation is plant3 = 2.95 + 0.0833 Water

Predictor       Coef       Stdev    t-ratio        p
Constant     2.95100     0.07797      37.85    0.000
Water        0.08334     0.01743       4.78    0.000

s = 0.2768      R-sq = 27.3%     R-sq(adj) = 26.1%
```

Regression of plant species 4 on water

```
The regression equation is plant4 = 1.38 - 0.0452 Water

Predictor       Coef       Stdev    t-ratio        p
Constant     1.37622     0.06129      22.45    0.000
Water       -0.04516     0.01371      -3.29    0.002

s = 0.2176      R-sq = 15.1%     R-sq(adj) = 13.7%
```



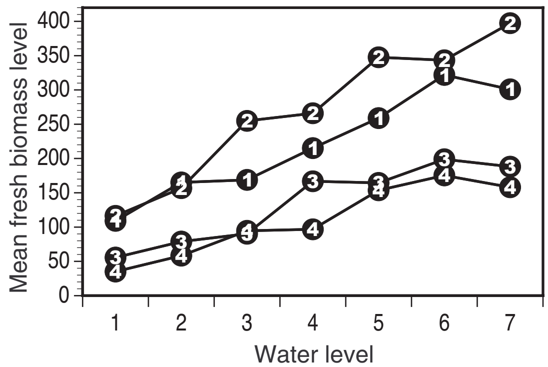
13.43. **(a)** & **(b)** The tables on the following page list the means and standard deviations (the latter in parentheses). The means plots show that biomass (both fresh and dry) increases with water level for all plants. Generally, plants 1 and 2 have higher biomass for each water level, while plants 3 and 4 are lower. Standard deviation ratios are quite high for both fresh and dry biomass: $108.01/6.79 \doteq 15.9$ and $35.76/3.12 \doteq 11.5$. **(c)** Minitab output below. For both fresh and dry biomass, main effects and the interaction are significant. (The interaction for fresh biomass has $P = 0.04$; other P-values are smaller.)

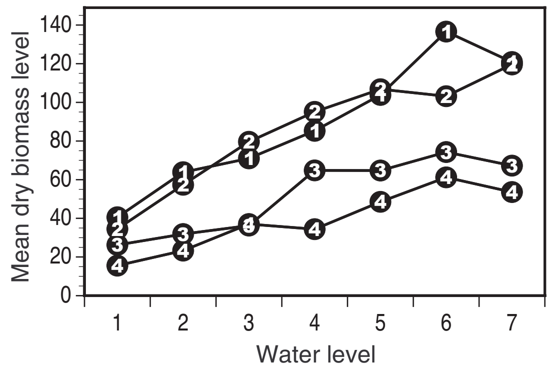
Minitab output: Two-way ANOVA for fresh biomass

Source	DF	SS	MS	F	P
Species	3	458295	152765	81.45	0.000
Water	6	491948	81991	43.71	0.000
Species*Water	18	60334	3352	1.79	0.040
Error	84	157551	1876		
Total	111	1168129			

Two-way ANOVA for dry biomass

Source	DF	SS	MS	F	P
Species	3	50523.8	16841.3	79.93	0.000
Water	6	56623.6	9437.3	44.79	0.000
Species*Water	18	8418.8	467.7	2.22	0.008
Error	84	17698.4	210.7		
Total	111	133264.6			





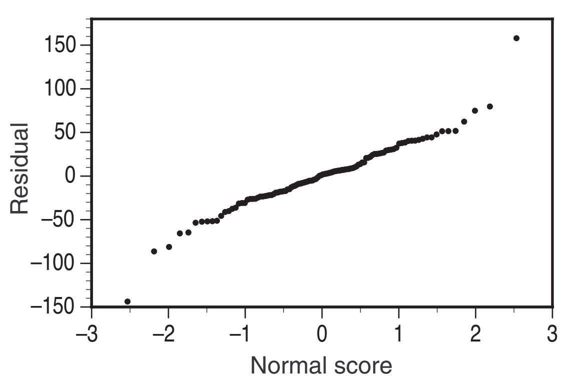
Fresh biomass

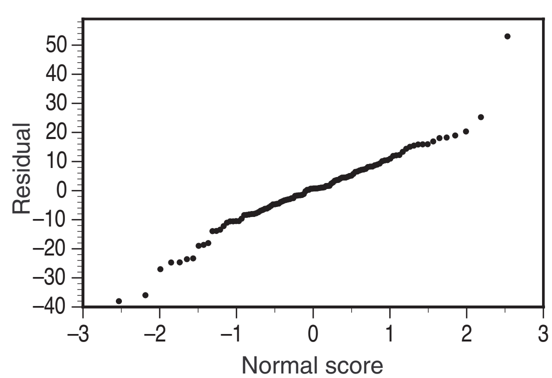
Species	50mm	150mm	250mm	350mm	450mm	550mm	650mm
1	109.095 (20.949)	165.138 (29.084)	168.825 (18.866)	215.133 (42.687)	258.900 (45.292)	321.875 (46.727)	300.880 (29.896)
2	116.398 (29.250)	156.750 (46.922)	254.875 (13.944)	265.995 (59.686)	347.628 (54.416)	343.263 (98.553)	397.365 (108.011)
3	55.600 (13.197)	78.858 (29.458)	90.300 (28.280)	166.785 (41.079)	164.425 (18.646)	198.910 (33.358)	188.138 (18.070)
4	35.128 (11.626)	58.325 (6.789)	94.543 (13.932)	96.740 (24.477)	153.648 (22.028)	175.360 (32.873)	158.048 (70.105)

Dry biomass

Species	50mm	150mm	250mm	350mm	450mm	550mm	650mm
1	40.565 (5.581)	63.863 (7.508)	71.003 (6.032)	85.280 (10.868)	103.850 (15.715)	136.615 (16.203)	120.860 (17.137)
2	34.495 (11.612)	57.365 (6.149)	79.603 (13.094)	95.098 (25.198)	106.813 (18.347)	103.180 (25.606)	119.625 (35.764)
3	26.245 (6.430)	31.865 (11.322)	36.238 (11.268)	64.800 (9.010)	64.740 (3.122)	74.285 (12.277)	67.258 (7.076)
4	15.530 (4.887)	23.290 (3.329)	37.050 (5.194)	34.390 (11.667)	48.538 (5.658)	61.195 (12.084)	53.600 (25.290)

13.44. Both sets of residuals have a high outlier (observation #53); observation #52 is a low outlier for fresh biomass. The other residuals look reasonably Normal.





13.45. For each water level, there is highly significant evidence of variation in biomass level (both fresh and dry) among plant species (Minitab output below). For each water level, we have df = 12, 6 comparisons, and $\alpha = 0.05$, so the Bonferroni critical value is $t^{**} = 3.1527$. (If we take into account that there are 7 water levels, so that overall we are performing $6 \times 7 = 42$ comparisons, we should take $t^{**} = 4.2192$.) The table below gives the pooled standard deviations s_p, the standard errors of each difference $\text{SE}_D = s_p\sqrt{1/4 + 1/4}$, and the "minimum significant difference" $\text{MSD} = t^{**}\text{SE}_D$ (two means are significantly different if they differ by at least this amount). MSD1 uses $t^{**} = 3.1527$, and MSD2 uses $t^{**} = 4.2192$. Rather than give a full listing of which differences are significant, we note that plants 3 and 4 are *not* significantly different, nor are 1 and 3 (except for one or two water levels). All other plant combinations are significantly different for at least three water levels. For fresh biomass, plants 2 and 4 are different for *all* levels, and for dry biomass, 1 and 4 differ for all levels.

	Fresh biomass				Dry biomass			
Water level	s_p	SE_D	MSD1	MSD2	s_p	SE_D	MSD1	MSD2
1	20.0236	14.1588	44.6382	50.3764	7.6028	5.3760	16.9487	19.1274
2	31.4699	22.2526	70.1552	79.1735	7.6395	5.4019	17.0305	19.2197
3	19.6482	13.8934	43.8012	49.4318	9.5103	6.7248	21.2010	23.9263
4	43.7929	30.9663	97.6265	110.1762	15.5751	11.0133	34.7213	39.1846
5	38.2275	27.0310	85.2197	96.1746	12.5034	8.8412	27.8734	31.4565
6	59.3497	41.9666	132.3068	149.3147	17.4280	12.3235	38.8518	43.8462
7	66.7111	47.1719	148.7174	167.8348	23.7824	16.8167	53.0176	59.8329

Minitab output: One-way ANOVA for fresh biomass — water level 1

```
Source    DF      SS      MS      F      p
Species    3   19107    6369  15.88  0.000
Error     12    4811     401
Total     15   23918
```

One-way ANOVA for fresh biomass — water level 2

```
Source    DF      SS      MS      F      p
Species    3   35100   11700  11.81  0.001
Error     12   11884     990
Total     15   46984
```

One-way ANOVA for fresh biomass — water level 3

```
Source    DF      SS      MS      F      p
Species    3   71898   23966  62.08  0.000
Error     12    4633     386
Total     15   76531
```

One-way ANOVA for fresh biomass — water level 4

```
Source    DF      SS      MS      F      p
Species    3   62337   20779  10.83  0.001
Error     12   23014    1918
Total     15   85351
```

One-way ANOVA for fresh biomass — water level 5

```
Source    DF      SS      MS      F      p
Species    3   99184   33061  22.62  0.000
Error     12   17536    1461
Total     15  116720
```

One-way ANOVA for fresh biomass — water level 6

```
Source    DF      SS      MS      F      p
Species    3   86628   28876   8.20  0.003
Error     12   42269    3522
Total     15  128897
```

One-way ANOVA for fresh biomass — water level 7

Source	DF	SS	MS	F	p
Species	3	144376	48125	10.81	0.001
Error	12	53404	4450		
Total	15	197780			

One-way ANOVA for dry biomass — water level 1

Source	DF	SS	MS	F	p
Species	3	1411.2	470.4	8.14	0.003
Error	12	693.6	57.8		
Total	15	2104.8			

One-way ANOVA for dry biomass — water level 2

Source	DF	SS	MS	F	p
Species	3	4597.1	1532.4	26.26	0.000
Error	12	700.3	58.4		
Total	15	5297.4			

One-way ANOVA for dry biomass — water level 3

Source	DF	SS	MS	F	p
Species	3	6127.2	2042.4	22.58	0.000
Error	12	1085.3	90.4		
Total	15	7212.6			

One-way ANOVA for dry biomass — water level 4

Source	DF	SS	MS	F	p
Species	3	8634	2878	11.86	0.001
Error	12	2911	243		
Total	15	11545			

One-way ANOVA for dry biomass — water level 5

Source	DF	SS	MS	F	p
Species	3	10026	3342	21.38	0.000
Error	12	1876	156		
Total	15	11902			

One-way ANOVA for dry biomass — water level 6

Source	DF	SS	MS	F	p
Species	3	13460	4487	14.77	0.000
Error	12	3645	304		
Total	15	17105			

One-way ANOVA for dry biomass — water level 7

Source	DF	SS	MS	F	p
Species	3	14687	4896	8.66	0.002
Error	12	6787	566		
Total	15	21474			

13.46. The F statistics for all eight ANOVAs are significant, and all eight regressions are significant as well. Unlike the nitrogen level (Exercises 13.39 through 13.42), all of these regressions have reasonably large values of R^2, and the scatterplots suggest that a straight line is an appropriate model for the relationship.

Minitab output: One-way ANOVA for fresh biomass — plant species 1

Source	DF	SS	MS	F	p
Water	6	145543	24257	19.76	0.000
Error	21	25774	1227		
Total	27	171317			

One-way ANOVA for fresh biomass — plant species 2

Source	DF	SS	MS	F	p
Water	6	257083	42847	9.63	0.000
Error	21	93463	4451		
Total	27	350546			

One-way ANOVA for fresh biomass — plant species 3

Source	DF	SS	MS	F	p
Water	6	80952	13492	17.77	0.000
Error	21	15948	759		
Total	27	96901			

One-way ANOVA for fresh biomass — plant species 4

Source	DF	SS	MS	F	p
Water	6	68704	11451	10.75	0.000
Error	21	22365	1065		
Total	27	91070			

One-way ANOVA for dry biomass — plant species 1

Source	DF	SS	MS	F	p
Water	6	27273	4545	30.44	0.000
Error	21	3136	149		
Total	27	30408			

One-way ANOVA for dry biomass — plant species 2

Source	DF	SS	MS	F	p
Water	6	21802	3634	7.83	0.000
Error	21	9751	464		
Total	27	31553			

One-way ANOVA for dry biomass — plant species 3

Source	DF	SS	MS	F	p
Water	6	9489.9	1581.6	18.82	0.000
Error	21	1764.6	84.0		
Total	27	11254.5			

One-way ANOVA for dry biomass — plant species 4

Source	DF	SS	MS	F	p
Water	6	6478	1080	7.44	0.000
Error	21	3047	145		
Total	27	9525			

Regression of fresh biomass+plant species 1 on water

```
The regression equation is plant1 = 80.1 + 35.0 Water

Predictor       Coef       Stdev    t-ratio        p
Constant       80.13       15.38       5.21    0.000
Water         34.961       3.438      10.17    0.000

s = 36.39       R-sq = 79.9%     R-sq(adj) = 79.1%
```

Regression of fresh biomass+plant species 2 on water

```
The regression equation is plant2 = 81.9 + 46.7 Water

Predictor       Coef       Stdev    t-ratio        p
Constant       81.94       26.97       3.04    0.005
Water         46.739       6.030       7.75    0.000

s = 63.82       R-sq = 69.8%     R-sq(adj) = 68.6%
```

Regression of fresh biomass+plant species 3 on water

```
The regression equation is plant3 = 33.0 + 25.4 Water

Predictor       Coef       Stdev    t-ratio        p
Constant       33.02       12.98       2.55    0.017
Water         25.423       2.901       8.76    0.000

s = 30.70       R-sq = 74.7%     R-sq(adj) = 73.7%
```

Regression of fresh biomass+plant species 4 on water

```
The regression equation is plant4 = 15.7 + 23.6 Water

Predictor       Coef       Stdev    t-ratio        p
Constant       15.69       13.98       1.12    0.272
Water         23.641       3.127       7.56    0.000

s = 33.09       R-sq = 68.7%     R-sq(adj) = 67.5%
```

Regression of dry biomass+plant species 1 on water

```
The regression equation is plant1 = 29.0 + 15.0 Water

Predictor       Coef       Stdev    t-ratio        p
Constant      28.971       6.033       4.80    0.000
Water         14.973       1.349      11.10    0.000

s = 14.28       R-sq = 82.6%     R-sq(adj) = 81.9%
```

Regression of dry biomass+plant species 2 on water

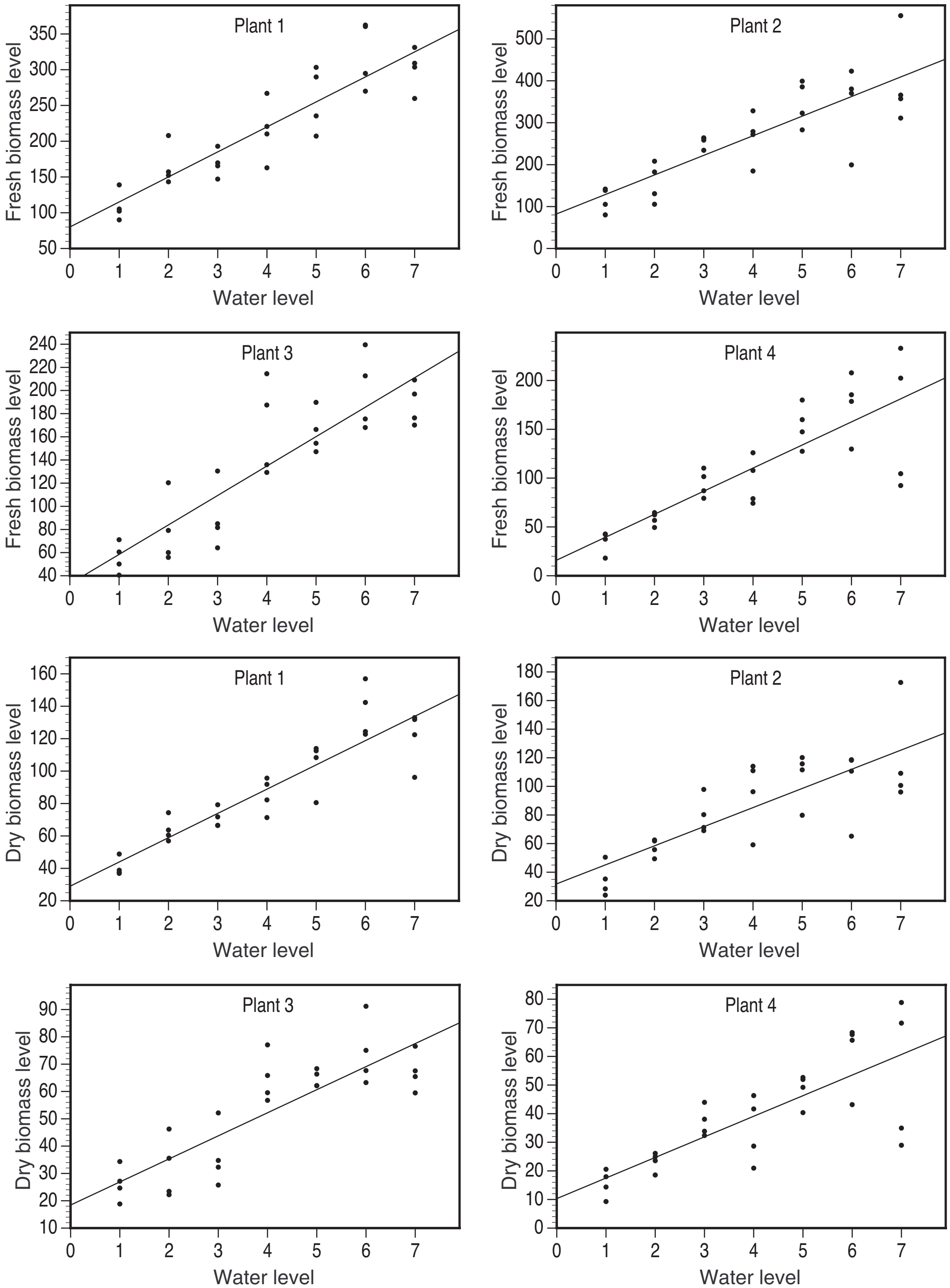
```
The regression equation is plant2 = 31.7 + 13.4 Water

Predictor       Coef       Stdev    t-ratio        p
Constant      31.707       8.905       3.56    0.001
Water         13.365       1.991       6.71    0.000

s = 21.07       R-sq = 63.4%     R-sq(adj) = 62.0%
```

Regression of dry biomass+plant species 3 on water

```
The regression equation is plant3 = 18.4 + 8.44 Water

Predictor       Coef       Stdev    t-ratio        p
Constant      18.436       4.741       3.89    0.001
Water          8.442       1.060       7.96    0.000

s = 11.22       R-sq = 70.9%     R-sq(adj) = 69.8%
```

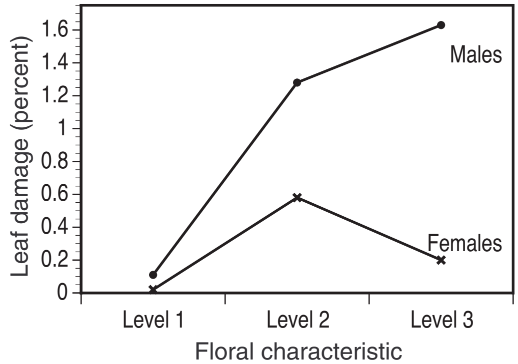
Regression of dry biomass+plant species 4 on water

```
The regression equation is plant4 = 10.3 + 7.20 Water

Predictor       Coef       Stdev    t-ratio        p
Constant      10.298       5.057       2.04    0.052
Water          7.197       1.131       6.36    0.000

s = 11.97       R-sq = 60.9%     R-sq(adj) = 59.4%
```


Plant 1
Fresh biomass level
Water level
Plant 2
Fresh biomass level
Water level
Plant 3
Fresh biomass level
Water level
Plant 4
Fresh biomass level
Water level
Plant 1
Dry biomass level
Water level
Plant 2
Dry biomass level
Water level
Plant 3
Dry biomass level
Water level
Plant 4
Dry biomass level
Water level

13.47. **(a)** With $I = 2$, $J = 3$, and $N = 180$, the numerator degrees of freedom are $I - 1$, $J - 1$, and $(I - 1)(J - 1)$, respectively, and the denominator degrees of freedom for all three tests is DFE $= N - IJ = 174$:

Source	df
Gender	1 and 174
Floral characteristic	2 and 174
Interaction	2 and 174



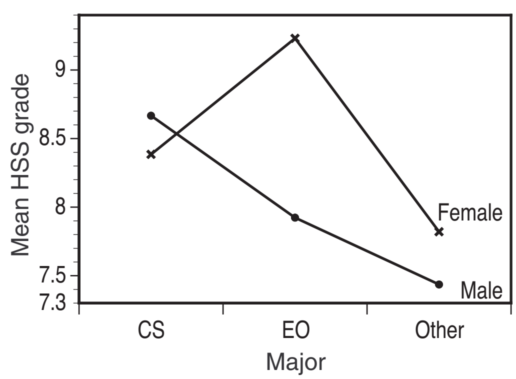
(b) Damage to males was higher for all characteristics. For males, damage was highest under characteristic level 3, while for females, the highest damage occurred at level 2. **(c)** Three of the standard deviations are at least half as large as the means. Because the response variable (leaf damage) had to be nonnegative, this suggests that these distributions are right-skewed; taking logarithms reduces the skewness.

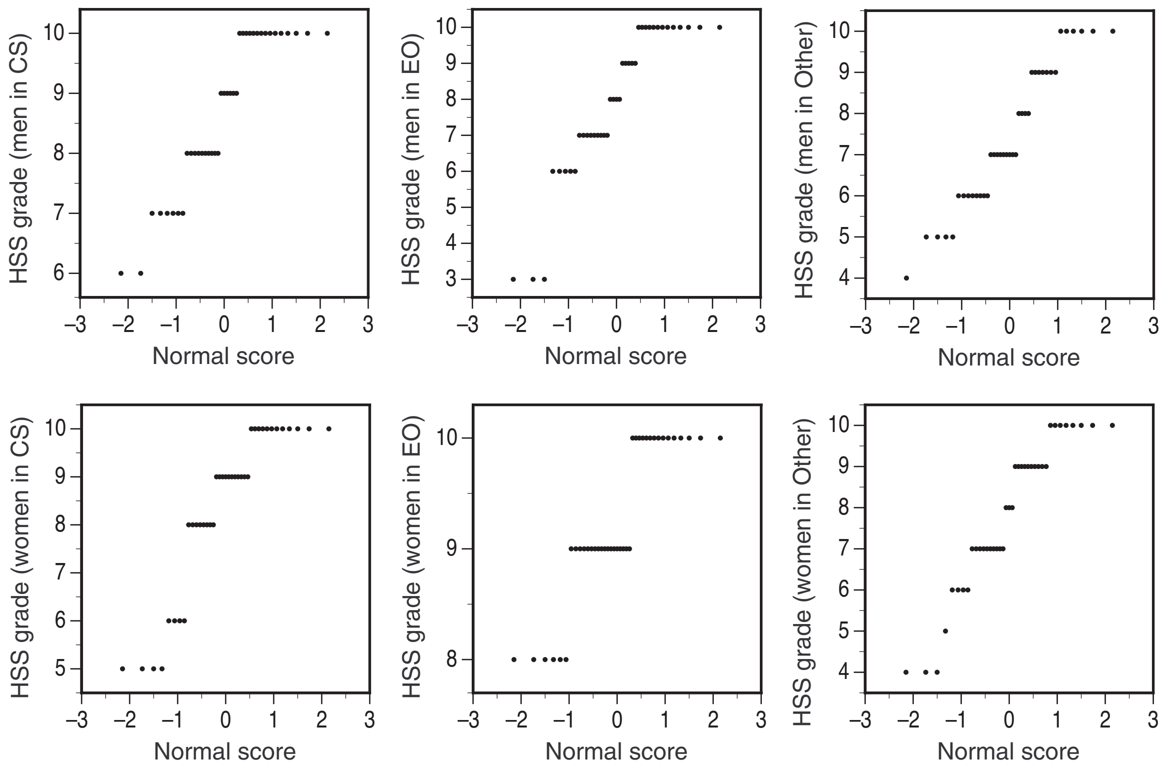
13.48. The table and plot of the means below suggest that, within a given gender, students who stay in the sciences have higher HSS grades than those who end up in the "Other" group. Males have a slightly higher mean in the CS group, but females have the edge in the other two. Normal quantile plots show no great deviations from Normality, apart from the granularity of the grades (most evident among women in EO). In the ANOVA, both main effects and the interaction are all significant. Residual analysis (not shown) shows that they are left-skewed.

Minitab output: Two-way ANOVA for HSS on sex and major

```
Source     DF        SS        MS      F      P
Sex         1    12.927    12.927   5.06  0.025
Maj         2    44.410    22.205   8.69  0.000
Sex*Maj     2    24.855    12.427   4.86  0.009
Error     228   582.923     2.557
Total     233   665.115
```

	Major		
Gender	CS	EO	Other
Male	$n = 39$	39	39
	$\bar{x} = 8.6667$	7.9231	7.4359
	$s = 1.2842$	2.0569	1.7136
Female	$n = 39$	39	39
	$\bar{x} = 8.3846$	9.2308	7.8205
	$s = 1.6641$	0.7057	1.8046



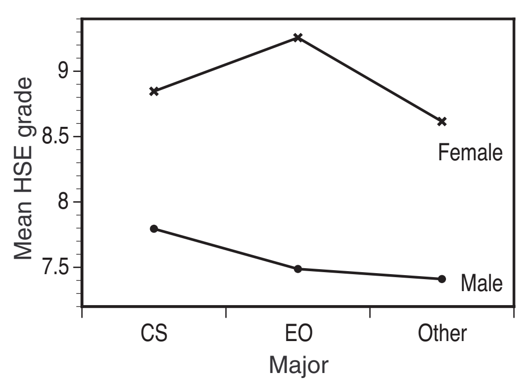


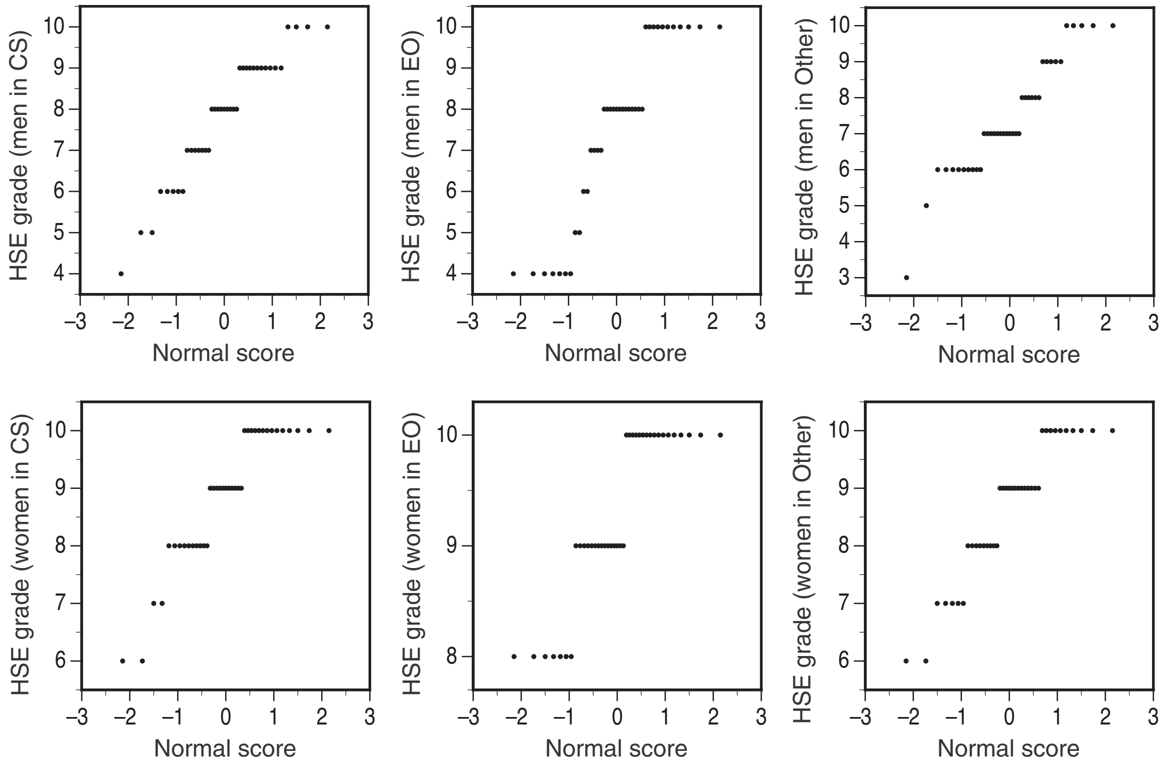
13.49. The table and plot of the means suggest that females have higher HSE grades than males. For a given gender, there is not too much difference among majors. Normal quantile plots show no great deviations from Normality, apart from the granularity of the grades (most evident among women in EO). In the ANOVA, only the effect of gender is significant. Residual analysis (not shown) reveals some causes for concern; for example, the variance does not appear to be constant.

Minitab output: Two-way ANOVA for HSE on sex and major

Source	DF	SS	MS	F	P
Sex	1	105.338	105.338	50.32	0.000
Maj	2	5.880	2.940	1.40	0.248
Sex*Maj	2	5.573	2.786	1.33	0.266
Error	228	477.282	2.093		
Total	233	594.073			

	Major		
Gender	CS	EO	Other
Male	$n = 39$	39	39
	$\bar{x} = 7.7949$	7.4872	7.4103
	$s = 1.5075$	2.1505	1.5681
Female	$n = 39$	39	39
	$\bar{x} = 8.8462$	9.2564	8.6154
	$s = 1.1364$	0.7511	1.1611



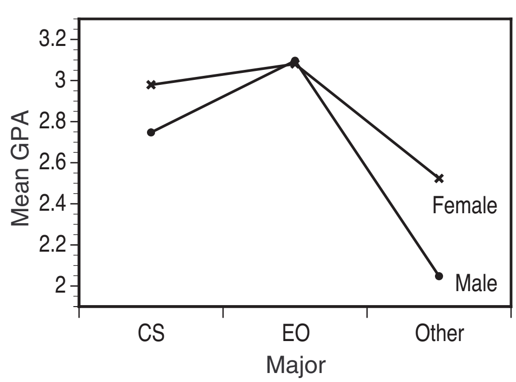


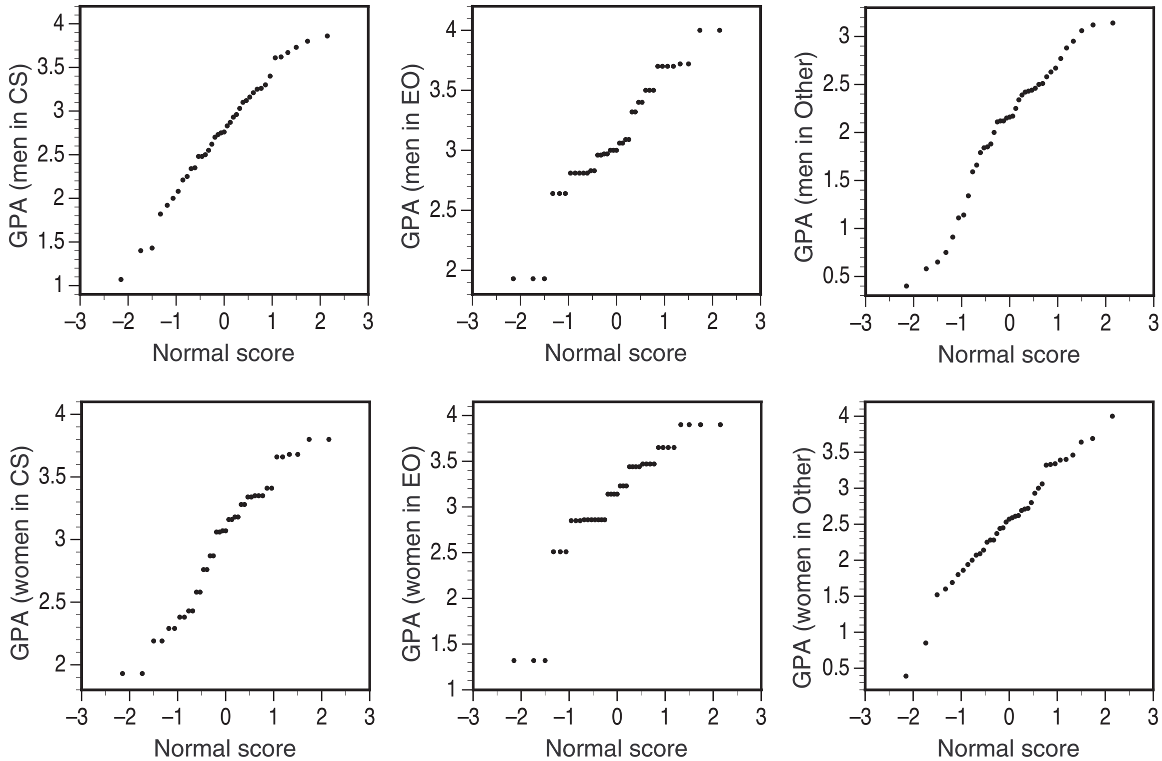
13.50. The table and plot of the means suggest that students who stay in the sciences have higher mean GPAs than those who end up in the "Other" group. Both genders have similar mean GPAs in the EO group, but in the other two groups, females perform better. Normal quantile plots show no great deviations from Normality, apart from a few low outliers in the two EO groups. In the ANOVA, sex and major are significant, while there is some (not quite significant) evidence for the interaction.

Minitab output: Two-way ANOVA for GPA on sex and major

```
Source      DF        SS        MS       F      P
Sex          1    3.1131    3.1131    7.31  0.007
Maj          2   26.7591   13.3795   31.42  0.000
Sex*Maj      2    2.3557    1.1779    2.77  0.065
Error      228   97.0986    0.4259
Total      233  129.3265
```

	Major		
Gender	CS	EO	Other
Male	$n = 39$	39	39
	$\bar{x} = 2.7474$	3.0964	2.0477
	$s = 0.6840$	0.5130	0.7304
Female	$n = 39$	39	39
	$\bar{x} = 2.9792$	3.0808	2.5236
	$s = 0.5335$	0.6481	0.7656



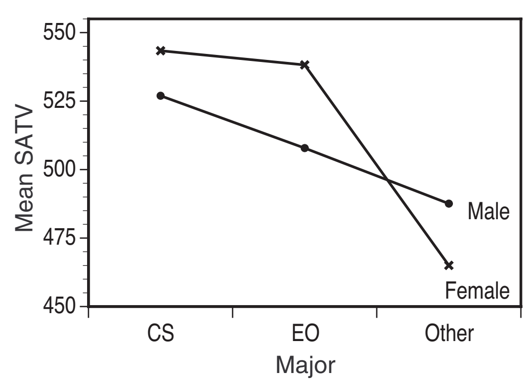


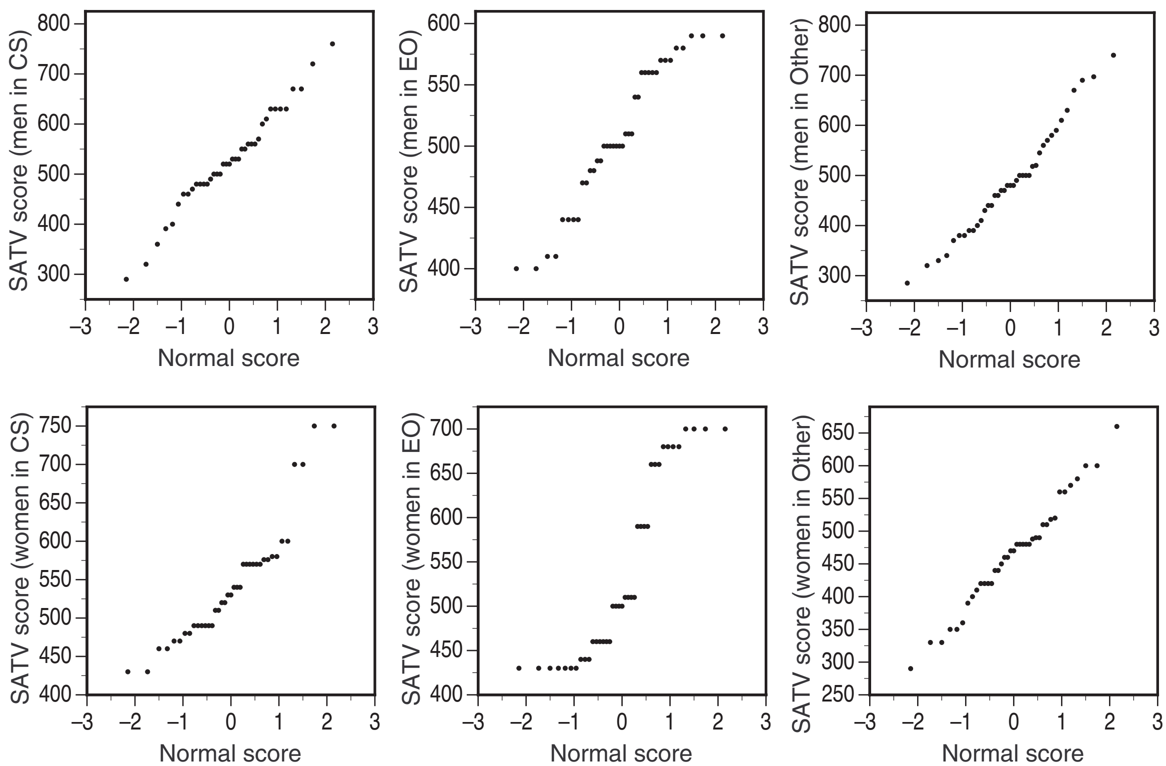
13.51. The table and plot of the means below suggest that students who stay in the sciences have higher mean SATV scores than those who end up in the "Other" group. Female CS and EO students have higher scores than males in those majors, but males have the higher mean in the Other group. Normal quantile plots suggest some right-skewness in the "Women in CS" group and also some non-Normality in the tails of the "Women in EO" group. Other groups look reasonably Normal. In the ANOVA table, only the effect of major is significant.

Minitab output: Two-way ANOVA for SATV on sex and major

Source	DF	SS	MS	F	P
Sex	1	3824	3824	0.47	0.492
Maj	2	150723	75362	9.32	0.000
Sex*Maj	2	29321	14661	1.81	0.166
Error	228	1843979	8088		
Total	233	2027848			

Gender		Major: CS	EO	Other
Male	$n =$	39	39	39
	$\bar{x} =$	526.949	507.846	487.564
	$s =$	100.937	57.213	108.779
Female	$n =$	39	39	39
	$\bar{x} =$	543.385	538.205	465.026
	$s =$	77.654	102.209	82.184




SATV score (men in CS)
Normal score
SATV score (men in EO)
Normal score
SATV score (men in Other)
Normal score
SATV score (women in CS)
Normal score
SATV score (women in EO)
Normal score
SATV score (women in Other)
Normal score

Chapter 14 Solutions

14.1. If $p = 0.5$, then odds $= \dfrac{p}{1-p} = \dfrac{0.5}{0.5} = 1$, or "1 to 1."

14.2. If odds $= 3$, then $p = \dfrac{\text{odds}}{\text{odds}+1} = \dfrac{3}{3+1} = \dfrac{3}{4}$.

14.3. We have $\hat{p}_{\text{men}} = \frac{63}{110} \doteq 0.5727$, and $\hat{p}_{\text{women}} = \frac{60}{130} = \frac{6}{13} \doteq 0.4615$. Therefore:

$$\text{odds}_{\text{men}} = \frac{63/110}{47/110} = \frac{63}{47} \doteq 1.3404, \text{ and}$$
$$\text{odds}_{\text{women}} = \frac{6/13}{7/13} = \frac{6}{7} \doteq 0.8571, \text{ or "6 to 7"}$$

Note: *The odds can also be computed without first finding $\hat{p}$; for example, 63 men preferred Commercial A and 47 preferred Commercial B, so odds$_{men} = \frac{63}{47}$.*

14.4. The odds for selecting Commercial B would be the reciprocal of the odds for Commercial A: $\text{odds}^*_{\text{men}} = \dfrac{47}{63} = 0.7460$ and $\text{odds}^*_{\text{women}} = \dfrac{7}{6} \doteq 1.1667$.

14.5. With $\text{odds}_{\text{men}} = \frac{63}{47}$ and $\text{odds}_{\text{women}} = \frac{6}{7}$, we have $\log(\text{odds}_{\text{men}}) \doteq 0.2930$ and $\log(\text{odds}_{\text{women}}) \doteq -0.1542$.

Note: *You may wish to remind students to use the natural logarithm, called "ln" by* Excel *and most calculators. A student who mistakenly uses the common (base-10) logarithm instead of the natural logarithm will get 0.1272 and −0.0669 as answers.*

14.6. With $\text{odds}^*_{\text{men}} = \frac{47}{63}$ and $\text{odds}^*_{\text{women}} = \frac{7}{6}$, we have $\log(\text{odds}^*_{\text{men}}) \doteq -0.2930$ and $\log(\text{odds}^*_{\text{women}}) \doteq 0.1542$.

Note: *Because these odds were the reciprocals of those from Exercise 14.3, the log odds are the opposites (negations) of those found in Exercise 14.5. A student who mistakenly uses the common (base-10) logarithm instead of the natural logarithm will get −0.1272 and 0.0669 as answers.*

14.7. The model is $y = \log(\text{odds}) = \beta_0 + \beta_1 x$. If $x = 1$ for men and 0 for women, we need:

$$\log\left(\frac{p_{\text{men}}}{1-p_{\text{men}}}\right) = \beta_0 + \beta_1 \quad \text{and} \quad \log\left(\frac{p_{\text{women}}}{1-p_{\text{women}}}\right) = \beta_0$$

We estimate $b_0 = \log(\text{odds}_{\text{women}}) \doteq -0.1542$ and $b_1 = \log(\text{odds}_{\text{men}}) - b_0 \doteq 0.4471$, so the regression equation is $\log(\text{odds}) = -0.1542 + 0.4471x$.

If $x = 0$ for men and 1 for women, we estimate $b_0 = \log(\text{odds}_{\text{men}}) \doteq 0.2930$ and $b_1 = \log(\text{odds}_{\text{women}}) - b_0 \doteq -0.4471$, so the regression equation is $\log(\text{odds}) = 0.2930 - 0.4471x$.

The estimated odds ratio is either:

$$e^{0.4471} \doteq \frac{\text{odds}_{\text{men}}}{\text{odds}_{\text{women}}} \doteq 1.5638 \quad \text{if } x = 1 \text{ for men, or}$$
$$e^{-0.4471} \doteq \frac{\text{odds}_{\text{women}}}{\text{odds}_{\text{men}}} \doteq 0.6395 \quad \text{if } x = 1 \text{ for women}$$

14.8. Because of the relationships between the (log) odds for selecting Commercial A and the (log) odds for selecting Commercial B, noted in the solutions to Exercises 14.4 and 14.6, these coefficients are the opposites (negations) of, and the odds ratios are reciprocals of, those found in the solution to the previous exercise.

The model is $y = \log(\text{odds}) = \beta_0 + \beta_1 x$. If $x = 1$ for men and 0 for women, we need:

$$\log\left(\frac{p_{\text{men}}}{1 - p_{\text{men}}}\right) = \beta_0 + \beta_1 \quad \text{and} \quad \log\left(\frac{p_{\text{women}}}{1 - p_{\text{women}}}\right) = \beta_0$$

We estimate $b_0 = \log(\text{odds}^*_{\text{women}}) \doteq 0.1542$ and $b_1 = \log(\text{odds}^*_{\text{men}}) - b_0 \doteq -0.4471$, so the regression equation is $\log(\text{odds}) = 0.1542 - 0.4471x$.

If $x = 0$ for men and 1 for women, we estimate $b_0 = \log(\text{odds}^*_{\text{men}}) \doteq -0.2930$ and $b_1 = \log(\text{odds}^*_{\text{women}}) - b_0 \doteq 0.4471$, so the regression equation is $\log(\text{odds}) = -0.293 + 0.4471x$.

The estimated odds ratio is either:

$$e^{-0.4471} \doteq \frac{\text{odds}_{\text{men}}}{\text{odds}_{\text{women}}} \doteq 0.6395 \quad \text{if } x = 1 \text{ for men, or}$$

$$e^{0.4471} \doteq \frac{\text{odds}_{\text{women}}}{\text{odds}_{\text{men}}} \doteq 1.5638 \quad \text{if } x = 1 \text{ for women}$$

14.9. (a) The appropriate test would be a chi-square test with df $= 5$. **(b)** The logistic regression model has no error term. **(c)** H_0 should refer to β_1 (the population slope) rather than b_1 (the estimated slope). **(d)** The interpretation of coefficients is affected by correlations among explanatory variables.

14.10. (a) $\beta_1 = 3$ means that log(odds) increases by 3 when x increases by 1. This means the odds increase by a factor of $e^3 \doteq 20$. **(b)** β_0 is the *log*-odds of an event. **(c)** The odds of an event is the ratio of the event's probability and its complement.

Note: *For part (a), it is difficult to make a simple statement about the effect on the probability when odds increases by a factor of 20. With a little algebra, we can start with the formula* $p = \frac{\text{odds}}{\text{odds} + 1}$ *and find that the new probability is* $p^* = \frac{20\text{odds}}{20\text{odds} + 1} = \frac{20}{19 + 1/p}$.

14.11. In each case, we compute

$$\log(\text{odds}) = -3.1658 + 1.3083x$$

and odds $= e^{\log(\text{odds})}$.

	$x =$ LOpening	log(odds)	odds
(a)	3.219	1.0456	2.8452
(b)	3.807	1.8149	6.1405
(c)	4.174	2.2950	9.9249

14.12. Use the formula given in Exercise 14.2: For each estimated odds value, the estimated probability is $\hat{p} = \frac{\text{odds}}{\text{odds} + 1}$. **(a)** $\frac{2.8452}{3.8452} \doteq 0.7399$. **(b)** $\frac{6.1405}{7.1405} \doteq 0.8600$. **(c)** $\frac{9.9249}{10.9249} \doteq 0.9085$.

14.13. (a) For each column, divide the "yes" entry by the total to find $\hat{p}$. **(b)** For each $\hat{p}$, compute odds $= \frac{\hat{p}}{1 - \hat{p}}$. **(c)** Finally, take log(odds).

$$\hat{p}_{\text{low}} = \tfrac{88}{1169} = 0.0753 \qquad \text{odds}_{\text{low}} \doteq 0.0814 \qquad \log(\text{odds}_{\text{low}}) \doteq -2.5083$$

$$\hat{p}_{\text{high}} = \tfrac{112}{1246} = 0.0899 \qquad \text{odds}_{\text{high}} \doteq 0.0988 \qquad \log(\text{odds}_{\text{high}}) \doteq -2.3150$$

14.14. **(a)** $\hat{p}_1 = \frac{108}{142} \doteq 0.7606$ for exclusive-territory firms. **(b)** $\hat{p}_2 = \frac{15}{28} \doteq 0.5357$ for other firms. **(c)** $\text{odds}_1 = \frac{\hat{p}_1}{1-\hat{p}_1} \doteq 3.1765$ and $\text{odds}_2 = \frac{\hat{p}_2}{1-\hat{p}_2} \doteq 1.1538$. **(d)** $\log(\text{odds}_1) \doteq 1.1558$ and $\log(\text{odds}_2) \doteq 0.1431$. (Be sure to use the *natural* logarithm for this computation.)

14.15. **(a)** $b_0 = \log(\text{odds}_{\text{low}}) \doteq -2.5083$ and $b_1 = \log(\text{odds}_{\text{high}}) - \log(\text{odds}_{\text{low}}) \doteq 0.1933$. **(b)** The fitted model is $\log(\text{odds}) = -2.5083 + 0.1933x$. **(c)** The odds ratio is $\text{odds}_{\text{high}}/\text{odds}_{\text{low}} = e^{b_1} \doteq 1.2132$ (or $\frac{0.0988}{0.0814} \doteq 1.2132$). **(d)** The relative risk from Example 9.7 was 1.19—very close to this odds ratio.

14.16. **(a)** $b_0 = \log(\text{odds}_2) \doteq 0.1431$ and $b_1 = \log(\text{odds}_1) - \log(\text{odds}_2) \doteq 1.0127$. **(b)** The fitted model is $\log(\text{odds}) = 0.1431 + 1.0127x$. **(c)** The odds ratio is $\text{odds}_1/\text{odds}_2 = e^{b_1} \doteq 2.7529$.

14.17. Shown below is Minitab output. **(a)** The slope is significantly different from 0 ($z = 2.37$, $P = 0.018$), but the constant is not ($z = 0.38$, $P = 0.706$). **(b)** With $b_1 = 1.0127$, $\text{SE}_{b_1} \doteq 0.4269$, and $z^* = 1.96$, the 95% confidence interval for β_1 is 0.176 to 1.849. **(c)** Exponentiating gives the interval $e^{0.176} \doteq 1.19$ to $e^{1.849} \doteq 6.36$.

Minitab output

```
Predictor      Coef    SE Coef     Z      P  Ratio  Lower  Upper
Constant   0.143101  0.378932  0.38  0.706
Exclusive
 Yes        1.01267  0.426920  2.37  0.018   2.75   1.19   6.36
```

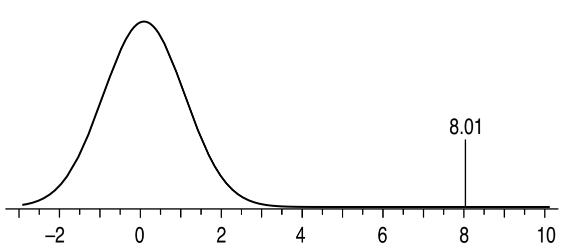
14.18. Recall that, by properties of exponents, $\frac{e^a}{e^b} = e^{a-b}$. Therefore:

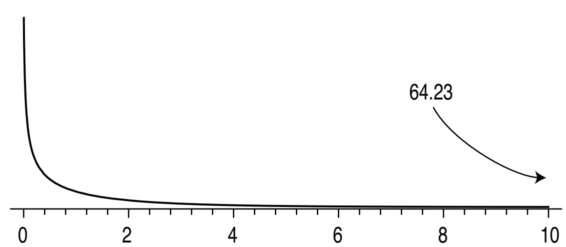
$$\frac{\text{odds}_{x+1}}{\text{odds}_x} = \frac{e^{-11.0391} \times e^{3.1709(x+1)}}{e^{-11.0391} \times e^{3.1709x}} = e^{3.1709(x+1)-3.1709x} = e^{3.1709(x+1-x)} = e^{3.1709}$$

14.19. With $b_1 \doteq 3.1088$ and $\text{SE}_{b_1} \doteq 0.3879$, the 99% confidence interval is $b_1 \pm 2.576\text{SE}_{b_1} \doteq b_1 \pm 0.9992$, or 2.1096 to 4.1080.

14.20. To find the confidence interval for the odds ratio, we first make a confidence interval for the slope b_1 and then transform (exponentiate) it: $b_1 \pm z^*\text{SE}_{b_1} = 3.1088 \pm (1.96)(0.3879) \doteq$ 2.3485 to 3.8691, so the odds ratio interval is $e^{2.3485} \doteq 10.470$ to $e^{3.8691} \doteq 47.898$. Up to rounding error, this agrees with the software output.

14.21. **(a)** $z = \frac{3.1088}{0.3879} \doteq 8.01$. **(b)** $z^2 \doteq 64.23$, which agrees with the value of X^2 given by SPSS and SAS. **(c)** The sketches are below. For both the Normal and chi-square distributions, the test statistics are quite extreme, consistent with the reported P-value.





14.22. Shown in the table below are the coefficients for the full model (from Example 14.11), as well as the three two-variable models and the three one-variable models (one of which appeared in Example 14.6). P-values for individual coefficients are given in parentheses below the coefficients. The X^2 statistics for the one-variable models are not shown; most software will not produce this, because the P-value for the coefficient measures the overall significance of the model.

	Coefficient of:					
Constant	LOpening	Theaters	Opinion	X^2	df	P
−2.0132	2.1467 (0.0277)	−0.0010 (0.2748)	−0.1095 (0.8083)	12.716	3	0.0053
−2.7164	2.1319 (0.0286)	−0.0010 (0.2805)		12.656	2	0.0018
−2.7154	1.3091 (0.0066)		−0.0710 (0.8672)	11.432	2	0.0033
−2.1815		0.00096 (0.0332)	−0.0065 (0.9858)	5.442	2	0.0658
−3.1658	1.3083 (0.0070)					
−2.2212		0.00096 (0.0329)				
−0.1230			0.0822 (0.8030)			

14.23. An odds ratio greater than 1 means a *higher* probability of a *low* tip. Therefore: The odds favor a low tip from senior adults, those dining on Sunday, those who speak English as a second language, and French-speaking Canadians. Diners who drink alcohol and lone males are less likely to leave low tips. For example, for a senior adult, the odds of leaving a low tip were 1.099 (for a probability of 0.5236).

14.24. **(a)** For each explanatory variable, we test H_0: $\beta_i = 0$ versus H_a: $\beta_i \neq 0$. **(b)** Under the null hypotheses, the X^2 statistic has a chi-square distribution with df $= 1$. Therefore, we reject H_0 at the 5% level if $X^2 > 3.84$. We do not reject H_0 for "Men's magazines" but have very strong evidence that all other coefficients (as well as the constant) are not zero. **(c)** The probability that the model's clothing is sexual is higher for magazines targeted at young adults (as the problem states), when the model is female, and for magazines aimed at men or at both men and women. **(d)** The fitted model is $\log(\text{odds}) = -2.32 + 0.50x_1 + 1.31x_2 - 0.05x_3 + 0.45x_4$.

14.25. **(a)** For men's magazines, the odds ratio confidence interval includes 1. This indicates that this explanatory variable has no effect on the probability that a model's clothing is not sexual, which is consistent with our failure to reject H_0 for men's magazines in the previous exercise. For all other explanatory variables, the odds ratio interval does not include 1, equivalent to the significant evidence against H_0 for those variables. **(b)** The odds that the model's clothing is not sexual are 1.27 to 2.16 times higher for magazines targeted at mature adults, 2.74 to 5.01 times higher when the model is male, and 1.11 to 2.23 times higher for magazines aimed at women. (These statements can also be made in terms of the odds that the model's clothing *is* sexual; for example, those odds are 1.27 to 2.16 times higher for magazines targeted at *young* adults, and so forth.) **(c)** The summary might note that it is easier to interpret the odds ratio rather than the regression coefficients because of the difficulty of thinking in terms of a log-odds scale.

14.26. (a) $\hat{p}_1 = \frac{463}{1000} = 0.463$. **(b)** $\text{odds}_1 = \frac{\hat{p}_1}{1-\hat{p}_1} \doteq 0.8622$. **(c)** $\hat{p}_2 = \frac{537}{1000} = 0.537$. **(d)** $\text{odds}_2 = \frac{\hat{p}_2}{1-\hat{p}_2} \doteq 1.1598$. **(e)** The odds in parts (b) and (d) are *reciprocals*—their product is 1. (Likewise, the probabilities in (a) and (c) are *complements*—their sum is 1.)

14.27. (a) $\hat{p}_{\text{hi}} = \frac{73}{91} \doteq 0.8022$ and $\text{odds}_{\text{hi}} = \frac{\hat{p}_{\text{hi}}}{1-\hat{p}_{\text{hi}}} = 4.0\overline{5}$. **(b)** $\hat{p}_{\text{non}} = \frac{75}{109} \doteq 0.6881$ and $\text{odds}_{\text{non}} = \frac{\hat{p}_{\text{non}}}{1-\hat{p}_{\text{non}}} \doteq 2.2059$. **(c)** The odds ratio is $\text{odds}_{\text{hi}}/\text{odds}_{\text{non}} \doteq 1.8385$. The odds of a high-tech company offering stock options are about 1.84 times those for a non-high-tech firm.

14.28. (a) $\log(\text{odds}_{\text{hi}}) \doteq 1.4001$ and $\log(\text{odds}_{\text{non}}) \doteq 0.7911$. **(b)** $\log(\text{odds}_{\text{non}}) = \beta_0$ and $\log(\text{odds}_{\text{hi}}) = \beta_0 + \beta_1$, so we find the estimates of β_0 and β_1 from the observed log-odds: $b_0 = \log(\text{odds}_{\text{non}}) \doteq 0.7911$ and $b_1 = \log(\text{odds}_{\text{hi}}) - \log(\text{odds}_{\text{non}}) \doteq 0.6090$. **(c)** $e^{b_1} \doteq e^{0.6090} \doteq 1.8385$, as we found in 16.13(c).

14.29. (a) With $b_1 \doteq 0.6090$ and $\text{SE}_{b_1} \doteq 0.3347$, the 95% confidence interval is $b_1 \pm 1.96\text{SE}_{b_1} \doteq b_1 \pm 0.6560$, or -0.0470 to 1.2650. **(b)** Exponentiating the confidence limits gives the interval 0.9540 to 3.5430. **(c)** Because the confidence interval for β_1 contains 0, or equivalently because 1 is in the interval for the odds ratio, we could not reject H_0: $\beta_1 = 0$ at the 5% level. There does not appear to be a significant difference between the odds of stock options for high-tech and other firms.

Note: *Software reports $z = 1.820$ and a P-value of 0.0688, which are nearly identical to the results for a two-proportion z test with the same counts ($z = -1.832$ and $P = 0.0669$)—see the solution to Exercise 8.67. For large samples, these two tests should give similar results.*

Minitab output: Logistic regression (high-tech versus non-high-tech companies)

```
                                               Odds     95% CI
Predictor      Coef     SE Coef     Z       P  Ratio   Lower  Upper
Constant   0.791128   0.206749   3.83   0.000
HT
 Yes       0.608960   0.334663   1.82   0.069   1.84    0.95   3.54
```

14.30. Minitab output is on the following page. All proportions, odds, odds ratios, and parameter estimates (b_0 and b_1) are unchanged. Because the standard error is smaller, the 95% confidence interval is narrower: $b_1 \pm 1.96\text{SE}_{b_1} \doteq b_1 \pm 0.4637$, or 0.1452 to 1.0727. The odds-ratio interval is therefore 1.1563 to 2.9233. Because 0 is not in the confidence interval for β_1 and 1 is not in the odds-ratio interval, we have significant evidence of a difference in the odds between the two types of companies.

Note: *For testing H_0: $\beta_1 = 0$, software reports $z = 2.573$ and $P = 0.0101$. For comparison, the test of $p_1 = p_2$ yields $z = 2.591$ and $P = 0.0096$.*

Minitab output: Logistic regression with sample sizes doubled

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI Lower	Upper
Constant	0.791128	0.146194	5.41	0.000			
HT							
Yes	0.608960	0.236642	2.57	0.010	1.84	1.16	2.92

14.31. **(a)** For the high blood pressure group, $\hat{p}_{\text{hi}} = \frac{55}{3338} \doteq 0.01648$, giving $\text{odds}_{\text{hi}} = \frac{\hat{p}_{\text{hi}}}{1-\hat{p}_{\text{hi}}} \doteq 0.01675$, or about 1 to 60. (If students give odds in the form "a to b," their choices of a and b might be different.) **(b)** For the low blood pressure group, $\hat{p}_{\text{lo}} = \frac{21}{2676} \doteq 0.00785$, giving $\text{odds}_{\text{lo}} = \frac{\hat{p}_{\text{lo}}}{1-\hat{p}_{\text{lo}}} \doteq 0.00791$, or about 1 to 126 (or 125). **(c)** The odds ratio is $\text{odds}_{\text{hi}}/\text{odds}_{\text{lo}} \doteq 2.1181$. Odds of death from cardiovascular disease are about 2.1 times greater in the high blood pressure group.

14.32. **(a)** For female references, $\hat{p}_{\text{w}} = \frac{48}{60} = 0.8$, giving $\text{odds}_{\text{w}} = \frac{\hat{p}_{\text{w}}}{1-\hat{p}_{\text{w}}} = 4$ ("4 to 1"). **(b)** For male references, $\hat{p}_{\text{m}} = \frac{52}{132} = 0.\overline{39}$, giving $\text{odds}_{\text{m}} = \frac{\hat{p}_{\text{m}}}{1-\hat{p}_{\text{m}}} = 0.65$ ("13 to 20"). **(c)** The odds ratio is $\text{odds}_{\text{w}}/\text{odds}_{\text{m}} \doteq 6.1538$. (The odds of a juvenile reference are more than six times greater for females.)

14.33. **(a)** The interval is $b_1 \pm 1.96\text{SE}_{b_1}$, or 0.2452 to 1.2558. **(b)** $X^2 = \left(\frac{0.7505}{0.2578}\right)^2 \doteq 8.47$. This gives a P-value between 0.0025 and 0.005. **(c)** We have strong evidence that there is a real (significant) difference in risk between the two groups.

14.34. **(a)** The interval is $b_1 \pm 1.96\text{SE}_{b_1}$, or 1.0946 to 2.5396. **(b)** $X^2 = \left(\frac{1.8171}{0.3686}\right)^2 \doteq 24.3$. This gives $P < 0.0005$. **(c)** We have strong evidence that there is a real (significant) difference in juvenile references between male and female references.

14.35. **(a)** The estimated odds ratio is $e^{b_1} \doteq 2.1181$ (as we found in Exercise 14.31). Exponentiating the interval for β_1 in Exercise 14.33(a) gives the odds-ratio interval from about 1.28 to 3.51. **(b)** We are 95% confident that the odds of death from cardiovascular disease are about 1.3 to 3.5 times greater in the high blood pressure group.

Minitab output: Logistic regression on blood pressure

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI Lower	Upper
Constant	-4.83968	0.219078	-22.09	0.000			
BP							
hi	0.750498	0.257840	2.91	0.004	2.12	1.28	3.51

14.36. **(a)** The estimated odds ratio is $e^{b_1} \doteq 6.1538$ (as we found in Exercise 14.32). Exponentiating the interval for β_1 in Exercise 14.34(a) gives the odds-ratio interval from about 2.99 to 12.67. **(b)** We are 95% confident that the odds of a juvenile reference are about 3 to 13 times greater among females.

Minitab output: Logistic regression on gender

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI Lower	Upper
Constant	-0.430783	0.178131	-2.42	0.016			
gender							
Female	1.81708	0.368641	4.93	0.000	6.15	2.99	12.67

14.37. **(a)** The model is $\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 x_i$, where $x_i = 1$ if the ith person is over 40, and 0 if he/she is under 40. **(b)** p_i is the probability that the ith person is terminated; this model assumes that the probability of termination depends on age (over/under 40). In this case, that seems to have been the case, but we might expect that other factors were taken into consideration. **(c)** The estimated odds ratio is $e^{b_1} \doteq 3.859$. (Of course, we can also get this from $\frac{41/765}{7/504}$.) We can also find, for example, a 95% confidence interval for b_1: $b_1 \pm 1.96\text{SE}_{b_1} = 0.5409$ to 2.1599. Exponentiating this translates to a 95% confidence interval for the odds: 1.7176 to 8.6701. The odds of being terminated are 1.7 to 8.7 times greater for those over 40. **(d)** Use a multiple logistic regression model, for example, $\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i}$.

14.38. **(a)** Positive coefficients indicate increasing odds (and increasing probability), and negative coefficients indicate decreasing odds. Therefore, the traits that make an individual more likely to use the Internet are those listed in the rightmost column of the table below. (The increase for having children is not significant.) **(b)** The odds ratios are given in the table below; for example, $e^{-0.063} \doteq 0.9389$ for Age. **(c)** The estimated log(odds) for this individual would be

$$-0.063(23) + 0.013(50) + 0.367(1) - 0.222(1) + 1.080(1) + 0.285(0) + 0.049(0) = 0.426$$

so the estimated odds would be $e^{0.426} \doteq 1.5311$. **(d)** The estimated probability is $p = \frac{\text{odds}}{\text{odds}+1} \doteq 0.6049$.

	b	odds ratio	Higher probability of Internet use
Age	−0.063	0.9389	younger
Income	0.013	1.0131	higher income
Location	0.367	1.4434	urban location
Sex	−0.222	0.8009	female
Education	1.080	2.9447	some post-secondary education
Language	0.285	1.3298	speak English
Children	0.049	1.0502	have children

14.39. It is difficult to find the needed probabilities from the numbers as given; this is made easier if we first convert the given information into a two-way table, shown on the right. The proportions meeting the activity guidelines are $\hat{p}_{\text{fruit}} = \frac{169}{237} \doteq 0.7131$ and $\hat{p}_{\text{no}} = \frac{494}{897} \doteq 0.5507$, so $\text{odds}_{\text{fruit}} \doteq 2.4853$ and $\text{odds}_{\text{no}} \doteq 1.2258$. Then $\log(\text{odds}_{\text{fruit}}) \doteq 0.9104$ and $\log(\text{odds}_{\text{no}}) \doteq 0.2036$, so $b_0 \doteq 0.2036$, $b_1 \doteq 0.7068$, and the model is $\log(\text{odds}) = 0.2036 + 0.7068x$. Software reports $\text{SE}_{b_1} \doteq 0.1585$ and $z \doteq 4.46$ for testing H_0: $\beta_1 = 0$. A 95% confidence interval for the odds ratio is 1.49 to 2.77.

	Eats fruit?		
Active?	Yes	No	Total
Yes	169	494	663
No	68	403	471
Total	237	897	1134

14.40. **(a)** For females, $\hat{p}_f = \frac{708}{1294} \doteq 0.5471$. For males, $\hat{p}_m = \frac{788}{1862} \doteq 0.4232$. **(b)** The odds for females and males are

$$\text{odds}_f = \frac{\hat{p}_f}{1-\hat{p}_f} \doteq 1.2082 \text{ and } \text{odds}_m = \frac{\hat{p}_m}{1-\hat{p}_m} \doteq 0.7337$$

so the odds ratio is $\frac{1.2082}{0.7337} \doteq 1.6467$. **(c)** The model is $\log(\text{odds}) = \beta_0 + \beta_1 x$, with $x = 0$ for male and $x = 1$ for female. (These values for x make the slope positive, because the odds are higher for females.) **(d)** $e^{b_1} \doteq 1.6467$, as we found in part (b). **(e)** The 95% confidence interval for β_1 is $b_1 \pm 1.96\,\text{SE}_{b_1} = 0.3559$ to 0.6417. Exponentiating gives a 95% confidence interval for the odds ratio: 1.4275 to 1.8997. Female odds of reducing spending are 1.4 to 1.9 times those of males.

14.41. For each group, the probability, odds, and log(odds) of being overweight are

$$\hat{p}_{no} = \frac{65080}{238215} \doteq 0.2732 \quad \text{odds}_{no} = \frac{\hat{p}_{no}}{1-\hat{p}_{no}} \doteq 0.3759 \quad \log(\text{odds}_{no}) \doteq -0.9785$$
$$\hat{p}_{FF} = \frac{83143}{291152} \doteq 0.2856 \quad \text{odds}_{FF} = \frac{\hat{p}_{FF}}{1-\hat{p}_{FF}} \doteq 0.3997 \quad \log(\text{odds}_{FF}) \doteq -0.9170$$

With $x = 0$ for no fast food and $x = 1$ for fast food, the logistic regression equation is $\log(\text{odds}) = -0.9785 + 0.0614x$. Software reports $\text{SE}_{b_1} \doteq 0.006163$, and for testing H_0: $\beta_1 = 0$ we have $z \doteq 9.97$, leaving little doubt that the slope is not 0. A 95% confidence interval for the odds ratio is 1.0506 to 1.0763; the odds of being overweight for students at schools close to fast-food restaurants are about 1.05 to 1.08 times greater than for students at schools that are not close to fast food.

14.42. **(a)** The researchers were adjusting for violations of independence: The samples could have included multiple students from the same school. **(b)** All of those other variables could have a connection to being overweight. If the researchers had not controlled for these variables, then their results might have been weakened (made less significant) if, for example, they had a slightly higher number of female students from rural schools, or a small number of non-exercising males in urban schools.

14.43. Portions of SAS and GLMStat output are given on the following page.
(a) The X^2 statistic for testing this hypothesis is 33.65 (df = 3), which has $P = 0.0001$. We conclude that at least one coefficient is not 0. **(b)** The fitted model is $\log(\text{odds}) = -6.053 + 0.3710\,\text{HSM} + 0.2489\,\text{HSS} + 0.03605\,\text{HSE}$. The standard errors of the three coefficients are 0.1302, 0.1275, and 0.1253, giving respective 95% confidence intervals 0.1158 to 0.6262, −0.0010 to 0.4988, and −0.2095 to 0.2816. **(c)** Only the coefficient of HSM is significantly different from 0, though HSS may also be useful.

Note: *In the multiple regression case study of Chapter 11, HSM was also the only significant explanatory variable among high school grades, and HSS was not even close to significant. See Figure 11.5 on page 603 of the text.*

SAS output

```
                              Intercept
                 Intercept       and
Criterion          Only      Covariates   Chi-Square for Covariates
-2 LOG L          295.340      261.691      33.648 with 3 DF (p=0.0001)

               Analysis of Maximum Likelihood Estimates

                 Parameter  Standard     Wald        Pr >     Standardized
Variable  DF      Estimate     Error  Chi-Square  Chi-Square     Estimate

INTERCPT   1      -6.0528    1.1562     27.4050      0.0001            .
HSM        1       0.3710    0.1302      8.1155      0.0044     0.335169
HSS        1       0.2489    0.1275      3.8100      0.0509     0.233265
HSE        1       0.0361    0.1253      0.0828      0.7736     0.029971
```

GLMStat output

```
             estimate   se(est)   z ratio   Prob>|z|
1 Constant  -6.053      1.156     -5.236     <0.0001
2 HSM        0.3710     0.1302     2.849      0.0044
3 HSS        0.2489     0.1275     1.952      0.0509
4 HSE         3.605e-2  0.1253     0.2877     0.7736
```

14.44. Portions of SAS and GLMStat output are given below. **(a)** The X^2 statistic for testing this hypothesis is 14.2 (df = 2), which has $P = 0.0008$. We conclude that at least one coefficient is not 0. **(b)** The model is $\log(\text{odds}) = -4.543 + 0.003690\,\text{SATM} + 0.003527\,\text{SATV}$. The standard errors of the two coefficients are 0.001913 and 0.001751, giving respective 95% confidence intervals −0.000059 to 0.007439, and 0.000095 to 0.006959. (The first coefficient has a P-value of 0.0537 and the second has $P = 0.0440$.) **(c)** We (barely) cannot reject $\beta_{\text{SATM}} = 0$—though because 0 is just in the confidence interval, we are reluctant to discard SATM. Meanwhile, we conclude that $\beta_{\text{SATV}} \neq 0$.

Note: *By contrast, with multiple regression of GPA on SAT scores, we found SATM useful but not SATV. See Figure 11.8 on page 607 of the text.*

SAS output

```
                              Intercept
                 Intercept       and
Criterion          Only      Covariates   Chi-Square for Covariates
-2 LOG L          295.340      281.119      14.220 with 2 DF (p=0.0008)

               Analysis of Maximum Likelihood Estimates

                 Parameter  Standard     Wald        Pr >     Standardized
Variable  DF      Estimate     Error  Chi-Square  Chi-Square     Estimate

INTERCPT   1      -4.5429    1.1618     15.2909      0.0001            .
SATM       1      0.00369   0.00191      3.7183      0.0538     0.175778
SATV       1      0.00353   0.00175      4.0535      0.0441     0.180087
```

GLMStat output

```
             estimate   se(est)   z ratio   Prob>|z|
1 Constant  -4.543      1.161     -3.915     <0.0001
2 SATM       3.690e-3   1.913e-3   1.929      0.0537
3 SATV       3.527e-3   1.751e-3   2.014      0.0440
```

14.45. The coefficients and standard errors for the fitted model are on the following page. Note that the tests requested in parts (a) and (b) are not available with all software packages.
(a) The X^2 statistic for testing this hypothesis is given by SAS as 19.2256 (df = 3); because $P = 0.0002$, we reject H_0 and conclude that high school grades add a significant amount

to the model with SAT scores. **(b)** The X^2 statistic for testing this hypothesis is 3.4635 (df $= 2$); because $P = 0.1770$, we cannot reject H_0; SAT scores do not add significantly to the model with high school grades. **(c)** For modeling the odds of HIGPA, high school grades (specifically HSM, and to a lesser extent HSS) are useful, while SAT scores are not.

SAS output

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Standardized Estimate
INTERCPT	1	-7.3732	1.4768	24.9257	0.0001	.
HSM	1	0.3427	0.1419	5.8344	0.0157	0.309668
HSS	1	0.2249	0.1286	3.0548	0.0805	0.210704
HSE	1	0.0190	0.1289	0.0217	0.8829	0.015784
SATM	1	0.000717	0.00220	0.1059	0.7448	0.034134
SATV	1	0.00289	0.00191	2.2796	0.1311	0.147566

Linear Hypotheses Testing

Label	Wald Chi-Square	DF	Pr > Chi-Square
HS	19.2256	3	0.0002
SAT	3.4635	2	0.1770

GLMStat output

	estimate	se(est)	z ratio	Prob>|z|
1 Constant	-7.373	1.477	-4.994	<0.0001
2 SATM	7.166e-4	2.201e-3	0.3255	0.7448
3 SATV	2.890e-3	1.914e-3	1.510	0.1311
4 HSM	0.3427	0.1419	2.416	0.0157
5 HSS	0.2249	0.1286	1.748	0.0805
6 HSE	1.899e-2	0.1289	0.1473	0.8829

14.46. **(a)** The fitted model is $\log(\text{odds}) = -0.6124 + 0.0609\ \text{Gender}$; the coefficient of gender is not significantly different from 0 ($z = 0.21$, $P = 0.8331$). **(b)** Now, $\log(\text{odds}) = -5.214 + 0.3028\ \text{Gender} + 0.004191\ \text{SATM} + 0.003447\ \text{SATV}$. In this model, gender is still not significant ($P = 0.3296$). **(c)** Gender is not useful for modeling the odds of HIGPA.

GLMStat output: Gender only

	estimate	se(est)	z ratio	Prob>|z|
1 Constant	-0.6124	0.4156	-1.474	0.1406
2 Gender	6.087e-2	0.2889	0.2107	0.8331

Gender and SAT scores

	estimate	se(est)	z ratio	Prob>|z|
1 Constant	-5.214	1.362	-3.828	0.0001
2 Gender	0.3028	0.3105	0.9750	0.3296
3 SATM	4.191e-3	1.987e-3	2.109	0.0349
4 SATV	3.447e-3	1.760e-3	1.958	0.0502

14.47. The models reported below are for the odds of *death*, as requested in the instructions. If a student models odds of survival, or codes the indicator variables for hospital and condition differently, his or her answers will be slightly different from these (but the conclusions should be the same). **(a)** The fitted model is $\log(\text{odds}) = -3.892 + 0.4157\ \text{Hospital}$, using 1 for Hospital A and 0 for Hospital B. With $b_1 \doteq -0.4157$ and $\text{SE}_{b_1} \doteq 0.2831$, we find that $z = -1.47$ or $X^2 = 2.16$ ($P = 0.1420$), so we do not have evidence to suggest that β_1 is not 0. A 95% confidence interval for β_1 is -0.1392 to 0.9706 (this interval includes 0). We estimate the odds ratio to be $e^{b_1} \doteq 1.515$, with confidence interval 0.87 to 2.64 (this includes 1, since β_1 might be 0). **(b)** The fitted model is $\log(\text{odds}) = -3.109 - 0.1320\,\text{Hospital} - 1.266\,\text{Condition}$; as before, use 1 for Hospital A and 0 for Hospital B, 1 for good condition and 0 for poor. The estimated odds ratio is $e^{b_1} \doteq 0.8764$, with confidence interval 0.48 to 1.60. **(c)** In neither case is the effect of Hospital significant. However, we can see the effect of Simpson's paradox in the coefficient of Hospital, or equivalently in the odds ratio. In the model with Hospital alone, this coefficient was positive and the odds ratio was greater than 1, meaning Hospital A patients have higher odds of death. When condition is added to the model, this coefficient is negative and the odds ratio is less than 1, meaning Hospital A patients have lower odds of death.

GLMStat output: Hospital only

```
             estimate   se(est)   z ratio   Prob>|z|
1 Constant   -3.892     0.2525    -15.41    <0.0001
2 Hosp        0.4157    0.2831    -1.469     0.1420

             odds ratio  lower 95% ci  upper 95% ci
1 Constant    2.041e-2    1.244e-2      3.348e-2
2 Hosp        1.515       0.8701        2.639
```

Hospital and condition

```
             estimate   se(est)   z ratio   Prob>|z|
1 Constant   -3.109     0.2959    -10.51    <0.0001
2 Hosp       -0.1320    0.3078    -0.4288    0.6681
3 Cond       -1.266     0.3218    -3.935    <0.0001

             odds ratio  lower 95% ci  upper 95% ci
1 Constant    4.463e-2    2.499e-2      7.971e-2
2 Hosp        0.8764      0.4794        1.602
3 Cond        0.2820      0.1501        0.5298
```

Chapter 15 Solutions

15.1. The rankings are shown on the right. Group A ranks (shaded) are 1, 2, 4, 6, and 8; Group B ranks are 3, 5, 7, 9, and 10.

Group	Rooms	Ranks
A	30	1
A	68	2
B	240	3
A	243	4
B	329	5
A	448	6
B	540	7
A	552	8
B	560	9
B	780	10

15.2. The list of ranks is not shown because it is nearly identical to the one shown in the previous solution; the only change needed is to change 780 to 4003 in the last line. The ranks assigned to each group are exactly the same.

15.3. The null hypothesis is H_0: no difference in distribution of number of rooms. The alternative might be two-sided ("there is a difference") or one-sided if we had a prior suspicion that one group had more rooms than the other. The test statistic is $W = 1 + 2 + 4 + 6 + 8 = 21$.

15.4. Changing the data does not change the hypotheses, so they are the same as in the previous solution. Additionally, because the assigned ranks did not change, the test statistic is still $W = 21$.

15.5. Under the null hypothesis,

$$\mu_W = \frac{(5)(11)}{2} = 27.5 \text{ and } \sigma_W = \sqrt{\frac{(5)(5)(11)}{12}} \doteq 4.7871$$

We found $W = 21$, so $z = \frac{21-27.5}{4.7871} \doteq -1.36$, for which the two-sided P-value is $2P(Z \le -1.36) \doteq 0.1738$. With the continuity correction, we find $z = \frac{21.5-27.5}{4.787} \doteq -1.25$, which gives $P = 2P(Z \le -1.25) = 0.2112$. The Minitab output on the following page gives a similar P-value to that found with the continuity correction; the difference is due to the rounding of z. Regardless of the P-value used, we do not reject H_0.

Note: *If a one-sided alternative was specified in Exercise 15.3, the P-value would be half as big: $P \doteq 0.0869$, or 0.1056 with the continuity correction.*

In the Minitab output, the medians are referred to as ETA1 and ETA2 ("eta" is the Greek letter η). Minitab also reports an estimate of 0.21 for the difference $\eta_1 - \eta_2$; note that this is not the same as the difference between the two sample medians ($0.7 - 0.4 = 0.3$). This estimate, called the Hodges-Lehmann estimate, is not discussed in the text and has been removed from the Minitab outputs accompanying other solutions for this chapter. Briefly, this estimate is found by taking every response from the first group and subtracting every response from the second group, yielding (in this case) a total of 25 differences. The median of this set of differences is the Hodges-Lehmann estimate.

Minitab output: Wilcoxon rank sum (Mann-Whitney) confidence interval and test

```
GrpA       N =   5     Median =       243.0
GrpB       N =   5     Median =       540.0
Point estimate for ETA1-ETA2 is      -228.0
96.3 Percent C.I. for ETA1-ETA2 is (-537.0,208.0)
W = 21.0
Test of ETA1 = ETA2  vs.  ETA1 ~= ETA2 is significant at 0.2101
```

15.6. Because the ranks and test statistic are unchanged, all answers are the same as those given in the previous solution.

15.7. (a) For example, $\frac{99}{254} \doteq 39.0\%$ of self-employed workers are completely satisfied. The complete table is on the right with a bar graph. Overall, self-employed workers are more satisfied than the other group. **(b)** See the Minitab output below: $X^2 = 15.641$ with df $= 3$, for which $P = 0.001$. We can reject H_0 and conclude that job satisfaction and job type (self-employed or not) are not independent.

Self-employed?				
Yes	39.0%	55.9%	3.1%	2.0%
No	28.2%	61.2%	8.2%	2.3%

Minitab output: Chi-square test

```
Expected counts are printed below observed counts

           C2       C3       C4       C5    Total
    1      99      142        8        5      254
        77.83   152.53    18.06     5.58

    2     250      542       73       20      885
       271.17   531.47    62.94    19.42

Total     349      684       81       25     1139

ChiSq =  5.760 +  0.727 +  5.606 +  0.059 +
         1.653 +  0.209 +  1.609 +  0.017 = 15.641
df = 3, p = 0.001
```

15.8. (a) Summary statistics for the two distributions are:

	$\bar{x}$	s	Min	Q_1	M	Q_3	Max
Men	15,287	7710	5180	9,951	12,423.5	21,791	29,920
Women	17,085	7858	7694	10,592	15,275.5	22,376	32,291

Men		Women
995	0	79
220	1	023
7	1	6
31	2	124
9	2	
	3	2

Various graphical summaries are possible; shown on the right is a back-to-back stemplot. The men's mean and median are lower than the women's, but the stemplots don't suggest a substantial difference. Neither distribution has extreme skewness or outliers. **(b)** The Wilcoxon test statistic is $W = 99$ with two-sided $P = 0.6776$ (Minitab output on the following page). We do not have enough evidence to conclude that there is difference between genders in words spoken.

Minitab output: Wilcoxon rank sum test

```
MWords     N =  10     Median =       12424
WWords     N =  10     Median =       15276
W = 99.0
Test of ETA1 = ETA2  vs.  ETA1 ~= ETA2 is significant at 0.6776
```

15.9. Back-to-back stemplots on the right, summary statistics below. The men's distribution is skewed, and the women's distribution has a near-outlier. Men and women are not significantly different ($W = 1421$, $P = 0.6890$). The t test assumes Normal distributions; with small samples (like the previous exercise), this might be risky. In this exercise, the samples might be large enough to overcome the apparent non-Normality of the distributions.

Men		Women
42210	0	2
99999876655	0	5557777888889
221100	1	01223444
77665	1	66666777789
4331	2	0112244
965	2	
1	3	2
6	3	

Note: *Shown below is the Minitab output for a t test; the conclusion is the same as the Wilcoxon test* ($t = -0.11$, $P = 0.92$).

	$\bar{x}$	s	Min	Q_1	M	Q_3	Max
Men	14,060	9065	695	7464.5	11118	22740	36345
Women	14,252	6515	2363	8345.5	14602	18050	32291

Minitab output: Wilcoxon rank sum test

```
Mwords     N =  37     Median =       11118
Wwords     N =  41     Median =       14602
W = 1421.0
Test of ETA1 = ETA2  vs.  ETA1 ~= ETA2 is significant at 0.6890
```

Two-sample *t* test

```
         N       Mean      StDev   SE Mean
Mwords  37      14060       9065      1490
Wwords  41      14252       6515      1017

T-Test mu Mwords = mu Wwords (vs not =): T= -0.11  P=0.92  DF=  64
```

15.10. **(a)** We find $W = 26$ and $P \doteq 0.0152$ (Minitab output on the following page). We have strong evidence against the hypothesis of identical distributions; we conclude that the weed-free yield is higher. **(b)** For testing H_0: $\mu_0 = \mu_9$ versus H_a: $\mu_0 > \mu_9$, we find $\bar{x}_0 = 170.2$, $s_0 \doteq 5.4216$, $\bar{x}_9 = 157.575$, $s_9 \doteq 10.1181$, and $t = 2.20$, which gives $P = 0.0423$ (df $= 4.6$). We have fairly strong evidence that the mean yield is higher with no weeds—but the evidence is not quite as strong as in (a). **(c)** Both tests still reach the same conclusion, so there is no "practically important impact" on our conclusions. The Wilcoxon evidence is slightly weaker: $W = 22$, $P \doteq 0.0259$. The t-test evidence is slightly stronger: $t = 2.79$, df $= 3$, $P = 0.0341$. The new statistics for the 9-weeds-per-meter group are $\bar{x}_9 = 162.633$ and $s_9 \doteq 0.2082$; these are substantial changes for each value.

Minitab output: Wilcoxon rank sum test (all points)

```
Weed0      N =   4     Median =      169.45
Weed9      N =   4     Median =      162.55
W = 26.0
Test of ETA1 = ETA2  vs.  ETA1 > ETA2 is significant at 0.0152
```

Wilcoxon rank sum test (outlier removed)

```
C2         N =   4     Median =      169.45
C4         N =   3     Median =      162.70
W = 22.0
Test of ETA1 = ETA2  vs.  ETA1 > ETA2 is significant at 0.0259
```

15.11. **(a)** Normal quantile plots are not shown. The score 0.00 for child 8 seems to be a low outlier (although with only five observations, such judgments are questionable). **(b)** For testing H_0: $\mu_1 = \mu_2$ versus H_a: $\mu_1 > \mu_2$, we have $\bar{x}_1 = 0.676$, $s_1 \doteq 0.1189$, $\bar{x}_2 = 0.406$, and $s_2 \doteq 0.2675$. Then, $t = 2.062$, which gives $P = 0.0447$ (df $= 5.5$). We have some evidence that high-progress readers have higher mean scores. **(c)** We test:

H_0: Scores for both groups are identically distributed
vs. H_a: High-progress children systematically score higher

for which we find $W = 36$ and $P \doteq 0.0473$ or 0.0463—significant evidence (at $\alpha = 0.05$) against the hypothesis of identical distributions. This is equivalent to the conclusion reached in part (b).

Minitab output: Wilcoxon rank sum test

```
HiProg1    N =   5     Median =      0.7000
LoProg1    N =   5     Median =      0.4000
W = 36.0
Test of ETA1 = ETA2  vs.  ETA1 > ETA2 is significant at 0.0473
The test is significant at 0.0463 (adjusted for ties)
```

15.12. **(a)** Normal quantile plots are not shown. The score 0.54 for child 3 seems to be a low outlier. **(b)** For testing H_0: $\mu_1 = \mu_2$ versus H_a: $\mu_1 > \mu_2$, we have $\bar{x}_1 = 0.768$, $s_1 \doteq 0.1333$, $\bar{x}_2 = 0.516$, $s_2 \doteq 0.2001$. Then, $t = 2.344$, which gives $P = 0.0259$ (df $= 6.97$). We have fairly strong evidence that high-progress readers have higher mean scores. **(c)** We test:

H_0: Scores for both groups are identically distributed
vs. H_a: High-progress children systematically score higher

for which we find $W = 38$ and $P \doteq 0.0184$. This is evidence against H_0, slightly stronger than that found in part (b).

Minitab output: Wilcoxon rank sum test

```
HiProg2    N =   5     Median =      0.8000
LoProg2    N =   5     Median =      0.4900
W = 38.0
Test of ETA1 = ETA2  vs.  ETA1 > ETA2 is significant at 0.0184
```

15.13. **(a)** See table. **(b)** For Story 2, $W = 8+9+4+7+10 = 38$. Under H_0:

$$\mu_W = \frac{(5)(11)}{2} = 27.5$$

$$\sigma_W = \sqrt{\frac{(5)(5)(11)}{12}} \doteq 4.7871$$

(c) $z = \frac{38-27.5}{4.787} \doteq 2.19$; with the continuity correction, we compute $\frac{37.5-27.5}{4.787} \doteq 2.09$, which gives $P = P(Z > 2.09) = 0.0183$. **(d)** See the table.

		Story 1		Story 2	
Child	Progress	Score	Rank	Score	Rank
1	high	0.55	4.5	0.80	8
2	high	0.57	6	0.82	9
3	high	0.72	8.5	0.54	4
4	high	0.70	7	0.79	7
5	high	0.84	10	0.89	10
6	low	0.40	3	0.77	6
7	low	0.72	8.5	0.49	3
8	low	0.00	1	0.66	5
9	low	0.36	2	0.28	1
10	low	0.55	4.5	0.38	2

15.14. **(a)** The outline is shown below. **(b)** We consider score improvements (posttest minus pretest). The means, medians, and standard deviations are:

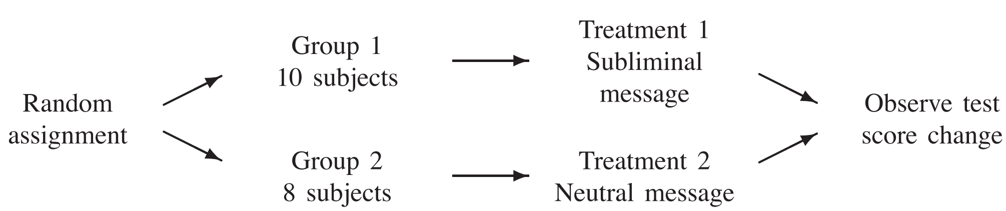
Treatment: $\bar{x} = 11.4$ $M = 11.5$ $s \doteq 3.1693$
Control: $\bar{x} = 8.25$ $M = 7.5$ $s \doteq 3.6936$

Treatment		Control
	0	455
76	0	7
	0	8
110	1	1
332	1	2
5	1	4
6	1	

A back-to-back stemplot is one way to compare the distributions graphically. Both of these comparisons support the idea that the positive subliminal message resulted in higher test scores. **(c)** We have $W = 114$, for which $P = 0.0501$ (or 0.0494, adjusted for ties). This is just about significant at $\alpha = 0.05$, and at least warrants further study.

Minitab output: Wilcoxon rank sum test

```
Trtmt      N =  10     Median =      11.500
Ctrl       N =   8     Median =       7.500
W = 114.0
Test of ETA1 = ETA2  vs.  ETA1 > ETA2 is significant at 0.0501
The test is significant at 0.0494 (adjusted for ties)
```



15.15. **(a)** At right. Unlogged plots appear to have a greater number of species. **(b)** We test H_0: There is no difference in the number of species on logged and unlogged plots versus H_a: Unlogged plots have a greater variety of species. The Wilcoxon test gives $W = 159$ and $P \doteq 0.0298$ (0.0290, adjusted for ties). We conclude that the observed difference is significant; unlogged plots really do have a greater number of species.

Unlogged		Logged
	0	4
	0	
	0	
	1	0
333	1	2
55	1	455
	1	7
998	1	88
10	2	
22	2	

Minitab output: Wilcoxon rank sum test

```
Unlogged   N =  12     Median =      18.500
Logged     N =   9     Median =      15.000
W = 159.0
Test of ETA1 = ETA2  vs.  ETA1 > ETA2 is significant at 0.0298
The test is significant at 0.0290 (adjusted for ties)
```

15.16. For the Wilcoxon test, we have $W = 579$, for which $P = 0.0064$ (0.0063, adjusted for ties). The evidence is slightly stronger with the Wilcoxon test than for the t and permutation tests.

Minitab output: Wilcoxon rank sum test

```
Trtmt      N =  21     Median =       53.00
Ctrl       N =  23     Median =       42.00
W = 579.0
Test of ETA1 = ETA2  vs.  ETA1 > ETA2 is significant at 0.0064
The test is significant at 0.0063 (adjusted for ties)
```

15.17. (a) We find $X^2 = 3.955$ with df $= (5-1)(2-1) = 4$, giving $P = 0.413$. There is little evidence to make us believe that there is a relationship between city and income.
(b) Minitab reports $W = 56{,}370$, with $P \doteq 0.5$; again, there is no evidence that incomes are systematically higher in one city.

Minitab output: Wilcoxon rank sum test

```
City1     N = 241     Median =      2.0000
City2     N = 218     Median =      2.0000
W = 56370.0
Test of ETA1 = ETA2  vs.  ETA1 ~= ETA2 is significant at 0.5080
The test is significant at 0.4949 (adjusted for ties)
```

15.18. We test:

$$H_0\text{: Food scores and activities scores have the same distribution}$$
$$\text{vs. } H_a\text{: Food scores are higher}$$

The differences, and their ranks, are:

Spa	Food score	Activities score	Difference	Rank
1	90.9	93.8	−2.9	4
2	92.3	92.3	0.0	–
3	88.6	91.4	−2.8	3
4	81.8	95.0	−13.2	6
5	85.7	89.2	−3.5	5
6	88.9	88.2	0.7	1
7	81.0	81.8	−0.8	2

In fact, it is not necessary to give the complete set of rankings; the only observations we need to make are (1) there is only one positive difference and (2) it is the smallest (in absolute value) of all the nonzero differences. Therefore, $W^+ = 1$.

Note: *In assigning ranks, differences of 0 are ignored; see the comment in the text toward the bottom of page 735. If a student mistakenly assigns a rank of 1 to 0, they would find $W^+ = 2$ (or perhaps 3 if they erroneously count 0 as a "positive difference").*

15.19. We test:

H_0: Food scores and activities scores have the same distribution
vs. H_a: Food scores are higher

The differences, and their ranks, are:

Spa	Food score	Activities score	Difference	Rank
1	77.3	95.7	−18.4	6
2	85.7	78.0	7.7	2
3	84.2	87.2	−3.0	1
4	85.3	85.3	0.0	–
5	83.7	93.6	−9.9	5
6	84.6	76.0	8.6	4
7	78.5	86.3	−7.8	3

The two positive differences have ranks 2 and 4, so $W^+ = 6$.

Note: *In assigning ranks, differences of 0 are ignored; see the comment in the text toward the bottom of page 735. If a student mistakenly assigns a rank of 1 to 0, they would find $W^+ = 3 + 5 = 8$ (or perhaps 9 if they erroneously count 0 as a "positive difference").*

15.20. Because one difference was 0, we ignore it and take $n = 6$, so that:

$$\mu_{W^+} = \frac{6(6+1)}{4} \doteq 10.5, \quad \sigma_{W^+} = \sqrt{\frac{6(6+1)(12+1)}{24}} = \sqrt{22.75} \doteq 4.7697,$$

and the approximate P-value is $P(W^+ \geq 0.5) \doteq P(Z \geq -2.10) \doteq 0.9821$. This agrees with the Minitab output below; see the note in the solution to Exercise 15.24 for an explanation of the estimated median reported by Minitab.

Note: *If a student does not see the instruction about discarding differences of 0 at the bottom of page 735, they might compute the mean and standard deviation using $n = 7$: $\mu_{W^+} = \frac{7(7+1)}{4} \doteq 14$ and $\sigma_{W^+} = \sqrt{\frac{7(7+1)(14+1)}{24}} = \sqrt{35} \doteq 5.9161$. Such a student would presumably take $W^+ = 2$ (or 3), so they would compute the approximate P-value as $P(W^+ \geq 1.5) \doteq P(Z \geq -2.11) \doteq 0.9826$ or $P(W^+ \geq 2.5) \doteq P(Z \geq -1.94) \doteq 0.9738$. While these are close to the right answer (and lead to the same conclusion), they are not quite correct. In other situations, failing to ignore differences of 0 may lead to the wrong conclusion.*

Minitab output: Wilcoxon signed rank rest (median = 0 versus median > 0)

```
                N FOR   WILCOXON            ESTIMATED
           N    TEST  STATISTIC  P-VALUE      MEDIAN
Diff       7       6        1.0    0.982      -2.000
```

15.21. Because one difference was 0, we ignore it and take $n = 6$, so that:

$$\mu_{W^+} = \frac{6(6+1)}{4} \doteq 10.5, \quad \sigma_{W^+} = \sqrt{\frac{6(6+1)(12+1)}{24}} = \sqrt{22.75} \doteq 4.7697,$$

and the approximate P-value is $P(W^+ \geq 5.5) \doteq P(Z \geq -1.05) \doteq 0.8531$. This agrees with the Minitab output below; see the note in the solution to Exercise 15.24 for an explanation of the estimated median reported by Minitab.

Note: *If a student does not see the instruction about discarding differences of 0 at the bottom of page 735, they might compute the mean and standard deviation using $n = 7$:*

$\mu_{W^+} = \frac{7(7+1)}{4} \doteq 14$ and $\sigma_{W^+} = \sqrt{\frac{7(7+1)(14+1)}{24}} = \sqrt{35} \doteq 5.9161$. Such a student would presumably take $W^+ = 8$ (or 9), so they would compute the approximate P-value as $P(W^+ \geq 7.5) \doteq P(Z \geq -1.10) \doteq 0.8643$ or $P(W^+ \geq 8.5) \doteq P(Z \geq -0.93) \doteq 0.8238$. While these are close to the right answer (and lead to the same conclusion), they are not quite correct. In other situations, failing to ignore differences of 0 may lead to the wrong conclusion.

Minitab output: Wilcoxon signed rank test (median = 0 versus median > 0)

```
                  N FOR   WILCOXON            ESTIMATED
             N     TEST  STATISTIC  P-VALUE      MEDIAN
Diff         7        6        6.0    0.853      -3.450
```

15.22. **(a)** Five of the six subjects rated drink A higher, by between 2 and 8 points. The subject who rated drink B higher only gave it a 2-point edge. **(b)** For testing H_0: $\mu_d = 0$ versus H_a: $\mu_d \neq 0$, we have $\bar{x} = 4.1\overline{6}$ and $s \doteq 3.6560$, so $t \doteq 2.79$ (df $= 5$) and $P \doteq 0.0384$—enough evidence to reject H_0. **(c)** For testing H_0: Ratings have the same distribution for both drinks versus H_a: One drink is systematically rated higher, we have $W^+ = 19.5$ and $P = 0.075$—not quite significant at $\alpha = 0.05$. **(d)** For a sample this small, the Wilcoxon test has low power. (See the related note in the solution to Exercise 15.24.)

Minitab output: Matched pairs *t* test

```
Variable     N       Mean     StDev   SE Mean         T     P-Value
Diff         6       4.17      3.66      1.49      2.79       0.038
```

Wilcoxon signed rank test

```
                  N FOR   WILCOXON            ESTIMATED
             N     TEST  STATISTIC  P-VALUE      MEDIAN
Diff         6        6       19.5    0.075       5.000
```

15.23. **(a)** With this additional subject, six of the seven subjects rated drink A higher, and (as before) the subject who preferred drink B only gave it a 2-point edge. **(b)** For testing H_0: $\mu_d = 0$ versus H_a: $\mu_d \neq 0$, we have $\bar{x} \doteq 7.8571$ and $s \doteq 10.3187$, so $t \doteq 2.01$ (df $= 6$) and $P \doteq 0.0906$. **(c)** For testing H_0: Ratings have the same distribution for both drinks versus H_a: One drink is systematically rated higher, we have $W^+ = 26.5$ and $P = 0.043$. **(d)** The new data point is an outlier (see the stemplot, above on the right), which may make the t procedure inappropriate. This also increases the standard deviation of the differences, which makes t insignificant. The Wilcoxon test is not sensitive to outliers, and the extra data point makes it powerful enough to reject H_0.

```
-0 | 2
 0 | 2
 0 | 5578
 1 |
 1 |
 2 |
 2 |
 3 | 0
```

Minitab output: Matched pairs *t* test

```
Variable     N       Mean     StDev   SE Mean         T     P-Value
Diff         7       7.86     10.32      3.90      2.01       0.091
```

Wilcoxon signed rank test

```
                  N FOR   WILCOXON            ESTIMATED
             N     TEST  STATISTIC  P-VALUE      MEDIAN
Diff         7        7       26.5    0.043       5.500
```

15.24. **(a)** The differences (treatment minus control) were 0.01622, 0.01102, and 0.01607. The mean difference was $\bar{x} \doteq 0.01444$, and $s \doteq 0.002960$. The fact that all are positive supports the idea that there was more growth in the treated plots. **(b)** For testing H_0: $\mu = 0$ versus H_a: $\mu > 0$, with μ the mean (treatment minus control) difference, we have $t = \frac{\bar{x}}{s/\sqrt{3}} \doteq 8.45$, df $= 2$, and $P = 0.0069$. We conclude that growth was greater in treated plots. **(c)** The

Wilcoxon statistic is $W^+ = 6$, for which $P = 0.091$. We would not reject H_0 (which states that there is no difference among pairs). **(d)** A low-power test has a low probability of rejecting H_0 when it is false.

Minitab output: Wilcoxon signed rank test (median = 0 versus median > 0)

```
             N FOR   WILCOXON           ESTIMATED
          N   TEST  STATISTIC  P-VALUE     MEDIAN
Diff      3      3        6.0    0.091    0.01485
```

Note: *With only three pairs, the Wilcoxon signed rank test can* never *give a P-value smaller than 0.091. This is one difference between some nonparametric tests and parametric tests like the t test: With the t test, the power improves when we consider alternatives that are farther from the null hypothesis; for example, if H_0 says $\mu = 0$, we have higher power for the alternative $\mu = 10$ than for $\mu = 5$. With the Wilcoxon signed rank test, all alternatives look the same; the values of W^+ and P would be the same if the three differences had been 100, 200, and 300.*

Also, note that the "estimated median" in the Minitab output (0.01485) is not the same as the median of the three differences (0.01607). The process of computing this point estimate is not discussed in the text, but we will illustrate it for this simple case: The Wilcoxon estimated median is the median of the set of Walsh averages *of the differences. This set consists of every possible pairwise average $(x_i + x_j)/2$ for $i \leq j$; note that this includes $i = j$, in which case the average is x_i. In general, there are $n(n+1)/2$ such averages, so with $n = 3$ differences, we have 6 Walsh averages: the three differences (0.01622, 0.01102, and 0.01607) and the averages of each pair of distinct differences (0.013545, 0.01362, and 0.016145). The median of 0.01102, 0.013545, 0.01362, 0.01607, 0.016145, and 0.01622 is 0.014845.*

15.25. We examine the heart-rate increase (final minus resting) from low-rate exercise; our hypotheses are H_0: median $= 0$ versus H_a: median > 0. The statistic is $W^+ = 10$ (the first four differences are positive, and the fifth is 0, so we drop it). We compute $P = P(W^+ \geq 9.5) = P\left(\frac{W^+ - 5}{2.739} \geq \frac{9.5 - 5}{2.739}\right) \doteq P(Z \geq 1.64) = 0.0505$. This is right on the borderline of significance: It is fairly strong evidence that heart rate increases, but (barely) not significant at 5%.

Minitab output: Wilcoxon signed rank test (median = 0 versus median > 0)

```
             N FOR   WILCOXON           ESTIMATED
          N   TEST  STATISTIC  P-VALUE     MEDIAN
LowDiff   5      4       10.0    0.050      7.500
```

15.26. (a) We first find the Final − Resting differences for both exercise rates (Low: 15, 9, 6, 9, 0; Medium: 21, 24, 15, 15, 18), then compute the differences of these differences (6, 15, 9, 6, 18). To this last list of differences, we apply the Wilcoxon signed rank test. The hypotheses are H_0: median $= 0$ versus H_a: median > 0. (The rank sum test is not appropriate because we do not have two independent samples.) **(b)** The statistic is $W^+ = 15$ (all five differences were positive), and the reported P-value is 0.030—fairly strong evidence that medium-rate exercise increases are greater than low-rate exercise increases.

Minitab output: Wilcoxon signed rank test (median = 0 versus median > 0)

```
             N FOR   WILCOXON           ESTIMATED
          N   TEST  STATISTIC  P-VALUE     MEDIAN
LowMed    5      5       15.0    0.030      10.50
```

15.27. For testing H_0: median $= 0$ versus H_a: median > 0, the Wilcoxon statistic is $W^+ = 119$ (14 of the 15 differences were positive, and the one negative difference was the smallest in absolute value), and $P < 0.0005$—very strong evidence that there are more aggressive incidents during moon days. This agrees with the results of the t and permutation tests. (See the note in the solution to Exercise 15.24 for an explanation of the estimated median reported by Minitab.)

Minitab output: Wilcoxon signed rank test (median = 0 versus median > 0)

```
                 N FOR   WILCOXON            ESTIMATED
            N    TEST   STATISTIC  P-VALUE     MEDIAN
diff       15      15       119.0    0.000      2.570
```

15.28. There are 17 nonzero differences; only one is negative (the boldface 6 in the list below).

Diff:	1	1	2	2	2	3	3	3	3	3	3	**6**	6	6	6	6	6
Rank:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Value:	1.5			4					8.5						14.5		

This gives $W^+ = 138.5$. (Note that the only tie we really need to worry about is the last group; all other ties involve only positive differences.)

15.29. (a) At right. The distribution is clearly right-skewed but has no outliers. **(b)** $W^+ = 31$ (only 4 of 12 differences were positive) and $P = 0.556$—there is no evidence that the median is other than 105. (See the note in the solution to Exercise 15.24 for an explanation of the estimated median reported by Minitab.)

9	1
9	5679
10	134
10	5
11	1
11	9
12	2

Minitab output: Wilcoxon signed rank test (median = 105 versus median ≠ 105)

```
                 N FOR   WILCOXON            ESTIMATED
            N    TEST   STATISTIC  P-VALUE     MEDIAN
Radon      12      12        31.0    0.556      103.2
```

15.30. If we compute Haiti content minus factory content (so that a negative difference means that the amount of vitamin C decreased), we find that the mean change is -5.33 and the median is -6. (See the note in the solution to Exercise 15.24 for an explanation of the estimated median reported by Minitab.) The stemplot is right-skewed. There are five positive differences; the Wilcoxon statistic is $W^+ = 37$, for which $P < 0.0005$. The differences are systematically negative, so vitamin C content is lower in Haiti.

−1	4
−1	3322
−1	
−0	9988
−0	7776666
−0	5444
−0	2
−0	1
0	1
0	33
0	4
0	
0	8

Minitab output: Wilcoxon signed rank test (median = 0 versus median < 0)

```
                 N FOR   WILCOXON            ESTIMATED
            N    TEST   STATISTIC  P-VALUE     MEDIAN
Change     27      27        37.0    0.000     -5.500
```

15.31. (a) The Wilcoxon statistic is $W^+ = 0$ (all of the differences were less than 16), for which $P = 0$—very strong evidence against H_0. We conclude that the median weight gain is less than 16 pounds. **(b)** Minitab (output below) gives the interval 3.75 to 5.90 kg for the median weight gain. (For comparison, in the solution to Exercise 7.32, the 95% confidence interval for the mean μ was about 3.80 to 5.66 kg. See the note in the solution to Exercise 15.24 for an explanation of the estimated median reported by Minitab.)

Minitab output: Wilcoxon signed rank test (median = 16 versus median ≠ 16)

```
                 N FOR   WILCOXON           ESTIMATED
            N     TEST  STATISTIC  P-VALUE     MEDIAN
Diff       16       16        0.0    0.000      4.800
```

Wilcoxon signed rank confidence interval

```
                 ESTIMATED    ACHIEVED
            N       MEDIAN  CONFIDENCE  CONFIDENCE INTERVAL
Diff       16         4.80        94.8  (    3.75,    5.90)
```

15.32. (a) For testing

H_0: The distribution of age at death is the same for all three groups
vs. H_a: At least one group is systematically higher or lower

we find $H = 11.11$ with df $= 2$, for which $P = 0.004$. **(b)** In the solution to Exercise 12.38, ANOVA yielded $F = 6.56$ (df 2 and 120) and $P = 0.002$. The conclusion is the same with either test.

Minitab output: Kruskal-Wallis test

```
LEVEL     NOBS    MEDIAN   AVE. RANK   Z VALUE
    1       67     73.00        63.4      0.46
    2       32     68.00        46.8     -2.81
    3       24     77.50        78.5      2.53
OVERALL    123                  62.0

H = 11.11  d.f. = 2  p = 0.004
H = 11.12  d.f. = 2  p = 0.004 (adjusted for ties)
```

15.33. (a) For testing

H_0: The distribution of BMD is the same for all three groups
vs. H_a: At least one group is systematically higher or lower

we find $H = 9.10$ with df $= 2$, for which $P = 0.011$. **(b)** In the solution to Exercise 12.39, ANOVA yielded $F = 7.72$ (df 2 and 42) and $P = 0.001$. The ANOVA evidence is slightly stronger, but (at $\alpha = 0.05$) the conclusion is the same.

Minitab output: Kruskal-Wallis test

```
LEVEL     NOBS    MEDIAN   AVE. RANK   Z VALUE
    1       15    0.2190        20.1     -1.05
    2       15    0.2160        17.7     -1.93
    3       15    0.2320        31.2      2.97
OVERALL     45                  23.0

H = 9.10  d.f. = 2  p = 0.011
H = 9.12  d.f. = 2  p = 0.011 (adjusted for ties)
```

15.34. **(a)** The Kruskal-Wallis test (Minitab output below) gives $H = 8.73$, df $= 4$, and $P = 0.069$—not significant at $\alpha = 0.05$. Note, however, that the rankings clearly suggest that vitamin C content decreases over time; the samples are simply too small to achieve significance even with such seemingly strong evidence. (See also a related comment in the solution to Exercise 15.24.) **(b)** The more accurate P-value is more in line with the apparent strength of the evidence, and does change our conclusion. With it, we reject H_0 and conclude that the distribution changes over time.

Minitab output: Kruskal-Wallis test

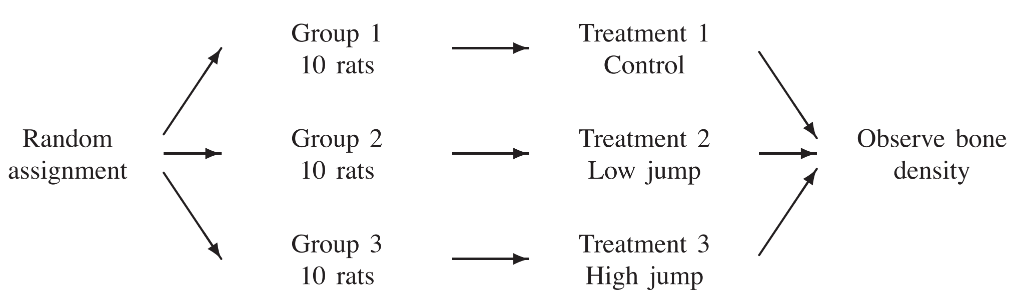
```
LEVEL     NOBS    MEDIAN  AVE. RANK   Z VALUE
    0        2    48.705        9.5      2.09
    1        2    41.955        7.5      1.04
    3        2    21.795        5.5      0.00
    5        2    12.415        3.5     -1.04
    7        2     8.320        1.5     -2.09
OVERALL     10                  5.5

H = 8.73  d.f. = 4  p = 0.069
```

15.35. **(a)** Diagram below. **(b)** The stemplots (right) suggest greater density for high-jump rats and a greater spread for the control group. **(c)** $H = 10.66$ with $P = 0.005$. We conclude that bone density differs among the groups. ANOVA tests H_0: all means are equal, assuming Normal distributions with the same standard deviation. For Kruskal-Wallis, the null hypothesis is that the distributions are the same (but not necessarily Normal). **(d)** There is strong evidence that the three groups have different bone densities; specifically, the high-jump group has the highest average rank (and the highest density), the low-jump group is in the middle, and the control group is lowest.

Control		Low jump		High jump	
55	4	55		55	
56	9	56		56	
57		57		57	
58		58	8	58	
59	33	59	469	59	
60	03	60	57	60	
61	14	61		61	
62	1	62		62	2266
63		63	1258	63	1
64		64		64	33
65	3	65		65	00
66		66		66	
67		67		67	4

Minitab output: Kruskal-Wallis test

```
LEVEL     NOBS    MEDIAN  AVE. RANK   Z VALUE
Ctrl        10     601.5       10.2     -2.33
Low         10     606.0       13.6     -0.81
High        10     637.0       22.6      3.15
OVERALL     30                 15.5

H = 10.66  d.f. = 2  p = 0.005
H = 10.68  d.f. = 2  p = 0.005 (adjusted for ties)
```



15.36. **(a)** For ANOVA, H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4$ versus H_a: Not all μ_i are equal. For Kruskal-Wallis, H_0 says that the distribution of the trapped insect count is the same for all board colors; the alternative is that the count is systematically higher for some colors. **(b)** In the order given, the medians are 46.5, 15.5, 34.5, and 15 insects; it appears that yellow is most effective, green is in the middle, and white and blue are least effective. The Kruskal-Wallis test statistic is $H = 16.95$, with df = 3; the P-value is 0.001, so we have strong evidence that color affects the insect count (that is, the difference we observed is statistically significant).

Minitab output: Kruskal-Wallis test

```
LEVEL     NOBS     MEDIAN  AVE. RANK   Z VALUE
Lemon        6      46.50       21.2      3.47
White        6      15.50        7.3     -2.07
Green        6      34.50       14.8      0.93
Blue         6      15.00        6.7     -2.33
OVERALL     24                  12.5

H = 16.95  d.f. = 3  p = 0.001
H = 16.98  d.f. = 3  p = 0.001 (adjusted for ties)
```

15.37. **(a)** $I = 4,\ n_i = 6,\ N = 24$. **(b)** The table below lists color, insect count, and rank. There are only three ties (and the second could be ignored, as both of those counts are for white boards). The R_i (rank sums) are:

$$\begin{array}{lllllll}
\text{Yellow} & 17 & +20 & +21 & +22 & +23 & +24 & = 127 \\
\text{White} & 3 & +\ 4 & +\ 5.5 & +\ 9.5 & +\ 9.5 & +12.5 & = 44 \\
\text{Green} & 7 & +14 & +15 & +16 & +18 & +19 & = 89 \\
\text{Blue} & 1 & +\ 2 & +\ 5.5 & +\ 8 & +11 & +12.5 & = 40
\end{array}$$

(c) $H = \dfrac{12}{24(25)}\left(\dfrac{127^2 + 44^2 + 89^2 + 40^2}{6}\right) - 3(25) = 91.95\overline{3} - 75 = 16.95\overline{3}$.
Under H_0, this has approximately the chi-squared distribution with df $= I - 1 = 3$; comparing to this distribution tells us that $0.0005 < P < 0.001$.

$$\begin{array}{cccccccccccccccccccccccc}
\text{B} & \text{B} & \text{W} & \text{W} & \text{W} & \text{B} & \text{G} & \text{B} & \text{W} & \text{W} & \text{B} & \text{W} & \text{B} & \text{G} & \text{G} & \text{G} & \text{Y} & \text{G} & \text{G} & \text{Y} & \text{Y} & \text{Y} & \text{Y} & \text{Y} \\
\hline
7 & 11 & 12 & 13 & 14 & 14 & 15 & 16 & 17 & 17 & 20 & 21 & 21 & 25 & 32 & 37 & 38 & 39 & 41 & 45 & 46 & 47 & 48 & 59 \\
\hline
1 & 2 & 3 & 4 & \multicolumn{2}{c}{\underbrace{5\quad 6}_{5.5}} & 7 & 8 & \multicolumn{2}{c}{\underbrace{9\quad 10}_{9.5}} & 11 & \multicolumn{2}{c}{\underbrace{12\quad 13}_{12.5}} & 14 & 15 & 16 & 17 & 18 & 19 & 20 & 21 & 22 & 23 & 24
\end{array}$$

15.38. **(a)** The stemplots (right) appear to suggest that logging reduces the number of species per plot and that recovery is slow (the 1-year-after and 8-years-after stemplots are similar). The logged stemplots have some outliers and appear to have more spread than the unlogged stemplot. The medians are 18.5, 12.5, and 15. **(b)** For testing H_0: all medians are equal versus H_a: at least one median is different, we have $H = 9.31$, df = 2, and $P = 0.010$ (or 0.009, adjusted for ties). This is good evidence of a difference among the groups.

Unlogged		1 year ago		8 years ago	
0		0	2	0	
0		0		0	4
0		0	7	0	
0		0	8	0	
1		1	11	1	0
1	333	1	23	1	2
1	55	1	4555	1	455
1		1		1	7
1	899	1	8	1	88
2	01	2		2	
2	22	2		2	

Minitab output: Kruskal-Wallis test for logging data

```
LEVEL     NOBS     MEDIAN  AVE. RANK   Z VALUE
    1       12      18.50       23.4      2.88
    2       12      12.50       11.5     -2.47
    3        9      15.00       15.8     -0.44
OVERALL     33                  17.0

H = 9.31  d.f. = 2  p = 0.010
H = 9.44  d.f. = 2  p = 0.009 (adjusted for ties)
```

15.39. **(a)** Yes, the data support this statement: The percent of high-SES subject who have never smoked ($\frac{68}{211} \doteq 32.2\%$) is higher than those percents for middle- and low-SES subjects (17.3% and 23.7%, respectively), and the percent of current smokers among high-SES subjects ($\frac{51}{211} \doteq 24.2\%$) is lower than among the middle- (42.3%) and low- (46.2%) SES groups. **(b)** $X^2 = 18.510$ with df $= 4$, for which $P = 0.001$. There is a significant relationship between SES and smoking behavior. **(c)** $H = 12.72$ with df $= 2$, so $P = 0.002$—or, after adjusting for ties, $H = 14.43$ and $P = 0.001$. The observed differences are significant; some SES groups smoke systematically more.

Minitab output: Chi-square test

```
          Never   Former     Curr    Total
 High        68       92       51      211
          58.68    83.57    68.75

  Mid         9       21       22       52
          14.46    20.60    16.94

  Low        22       28       43       93
          25.86    36.83    30.30

Total        99      141      116      356

ChiSq =   1.481 +  0.850 +  4.584 +
          2.062 +  0.008 +  1.509 +
          0.577 +  2.119 +  5.320 = 18.510
df = 4, p = 0.001
```

Kruskal-Wallis test

```
LEVEL   NOBS   MEDIAN  AVE. RANK  Z VALUE
 High    211    2.000      162.4    -3.56
  Mid     52    2.000      203.6     1.90
  Low     93    2.000      201.0     2.46
OVERALL  356               178.5

H = 12.72  d.f. = 2  p = 0.002
H = 14.43  d.f. = 2  p = 0.001
                       (adjusted for ties)
```

15.40. **(a)** Choice of graphical summaries will vary; a back-to-back stemplot is shown on the right.

Women		Men
	0	033334
96	0	66679999
22222221	1	2222222
888888888875555	1	558
4440	2	00344
	2	
	3	0
6	3	

	$\bar{x}$	s	Median
Women	$165.1\overline{6}$	56.515	175
Men	$117.1\overline{6}$	74.240	120

(b) The Wilcoxon rank sum test yields $W = 1105.5$, with two-sided $P = 0.0050$—significant evidence of a difference. (Minitab output on the following page.) **(c)** The t test yields $t \doteq 2.82$ with df $\doteq 54.2$ and $P \doteq 0.0067$. **(d)** Both distributions have a high outlier, and the men's distribution is skewed, making the use of a t test somewhat risky.

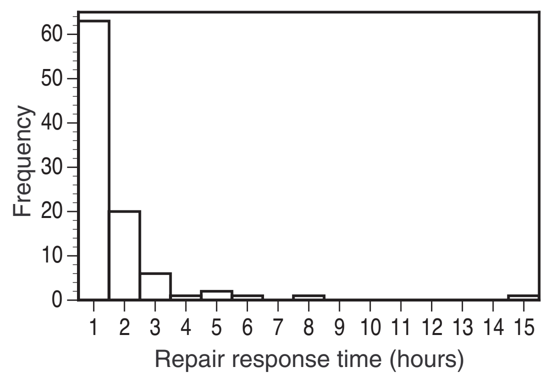
Minitab output: Wilcoxon rank sum test

```
studyF      N =  30     Median =      175.00
studyM      N =  30     Median =      120.00
W = 1105.5
Test of ETA1 = ETA2  vs.  ETA1 ~= ETA2 is significant at 0.0050
The test is significant at 0.0045 (adjusted for ties)
```

Two-sample *t* test

```
        N      Mean     StDev   SE Mean
studyF  30     165.2     56.5        10
studyM  30     117.2     74.2        14

95% C.I. for mu study - mu C8: ( 14,  82)
T-Test mu study = mu C8 (vs not =): T= 2.82  P=0.0067  DF=  54
```

15.41. **(a)** On the right is a histogram of service times for Verizon customers. With only 10 CLEC service calls, it is hardly necessary to make such a graph for them; we can simply observe that 7 of those 10 calls took 5 hours, which is quite different from the distribution for Verizon customers. The means and medians tell the same story:



Verizon	$\bar{x}_V \doteq 1.7263$ hr	$M_V = 1$ hr
CLEC	$\bar{x}_C = 3.8$ hr	$M_C = 5$ hr

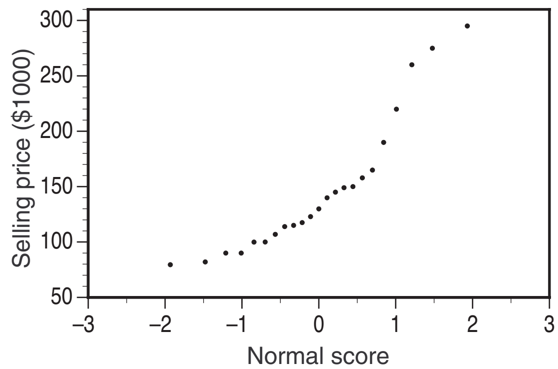
(b) The distributions are sharply skewed, and the sample sizes are quite different; the t test is not reliable in situations like this. The Wilcoxon rank-sum test gives $W = 4778.5$, which is highly significant ($P = 0.0026$ or 0.0006). We have strong evidence that response times for Verizon customers are shorter. It is also possible to apply the Kruskal-Wallis test (with two groups). While the P-values are slightly different ($P = 0.005$, or 0.001 adjusted for ties), the conclusion is the same: We have strong evidence of a difference in response times.

Minitab output: Wilcoxon rank sum test

```
Verizon     N =  95     Median =       1.000
CLEC        N =  10     Median =       5.000
W = 4778.5
Test of ETA1 = ETA2  vs.  ETA1 < ETA2 is significant at 0.0026
The test is significant at 0.0006 (adjusted for ties)
```

Kruskal-Wallis test

```
LEVEL    NOBS    MEDIAN  AVE. RANK   Z VALUE
    1      95     1.000       50.3     -2.80
    2      10     5.000       78.7      2.80
OVERALL   105                 53.0

H =  7.84  d.f. = 1  p = 0.005
H = 10.54  d.f. = 1  p = 0.001 (adjusted for ties)
```



15.42. Stemplots and other details can be found in the solution to Exercise 7.141. **(a)** The distribution of prices for three-bedroom houses is clearly right-skewed, with high outliers. **(b)** For testing H_0: $\mu_3 = \mu_4$ versus H_a: $\mu_3 \neq \mu_4$, we have $t \doteq -4.475$ with either df $= 20.98$ ($P \doteq 0.0002$) or df $= 13$ ($P < 0.001$). We reject H_0 and conclude that the mean prices are different (specifically, that 4BR houses are more expensive). **(c)** We use the Wilcoxon rank sum test for the hypotheses H_0: medians are equal versus H_a: medians are different. We find $W = 312$ and $P \doteq 0.0001$—significant evidence that prices differ. This is equivalent to the conclusion reached in part (b).

Minitab output: Two-sample *t* test

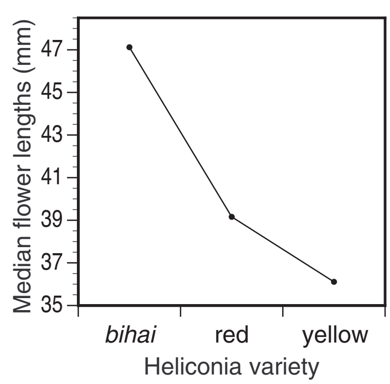
```
      N      Mean     StDev   SE Mean
3BR  23    147561     61741     12874
4BR  14    266793     87275     23325

95% C.I. for mu 3BR - mu 4BR: ( -174820,  -63644)
T-Test mu 3BR = mu 4BR (vs not =): T= -4.48  P=0.0002  DF=  20
```

Wilcoxon rank sum test

```
3BR        N =  23     Median =      129900
4BR        N =  14     Median =      259900
W = 312.0
Test of ETA1 = ETA2  vs.  ETA1 ~= ETA2 is significant at 0.0001
The test is significant at 0.0001 (adjusted for ties)
```

15.43. See also the solutions to Exercises 1.79 and 12.35; the latter exercise requests the same analysis for ANOVA. The means, standard deviations, and medians (all in millimeters) are:

Variety	n	$\bar{x}$	s	M
bihai	16	47.5975	1.2129	47.12
red	23	39.7113	1.7988	39.16
yellow	15	36.1800	0.9753	36.11



The appropriate rank test is a Kruskal-Wallis test of H_0: all three varieties have the same length distribution versus H_a: at least one variety is systematically longer or shorter. We reject H_0 and conclude that at least one species has different lengths ($H = 45.35$, df $= 2$, $P < 0.0005$).

Minitab output: Kruskal-Wallis test

```
LEVEL     NOBS     MEDIAN  AVE. RANK   Z VALUE
    1       16      47.12       46.5      5.76
    2       23      39.16       26.7     -0.32
    3       15      36.11        8.5     -5.51
OVERALL     54                  27.5

H = 45.35  d.f. = 2  p = 0.000
H = 45.36  d.f. = 2  p = 0.000 (adjusted for ties)
```

15.44. **(a)** The mean and median suggest that iron content is least for aluminum pots and greatest for iron pots. ANOVA requires Normal data with equal standard deviations; the former is difficult to assess with such small samples, and for the latter, the largest-to-smallest ratio is 1.99—just within our guidelines for pooling. **(b)** The Kruskal-Wallis test gives $H = 8.00$, df $= 2$, $P = 0.019$. We conclude that vegetable iron content differs by pot type.

	n	M	$\bar{x}$	s
Aluminum	4	1.185	1.2325	0.2313
Clay	4	1.615	1.46	0.4601
Iron	4	2.79	2.79	0.2399

Minitab output: Kruskal-Wallis test for vegetable dish

```
LEVEL    NOBS    MEDIAN  AVE. RANK   Z VALUE
 Alum       4     1.185        3.5     -2.04
 Clay       4     1.615        5.5     -0.68
 Iron       4     2.860       10.5      2.72
OVERALL    12                  6.5

H = 8.00  d.f. = 2  p = 0.019
```

15.45. Use the Wilcoxon rank sum test with a two-sided alternative. For meat, $W = 15$ and $P = 0.4705$, and for legumes, $W = 10.5$ and $P = 0.0433$ (or 0.0421). There is no evidence of a difference in iron content for meat, but for legumes the evidence is significant at $\alpha = 0.05$.

Minitab output: Wilcoxon rank sum test for meat

```
Alum       N =   4     Median =       2.050
Clay       N =   4     Median =       2.375
W = 15.0
Test of ETA1 = ETA2  vs.  ETA1 ~= ETA2 is significant at 0.4705
Cannot reject at alpha = 0.05
```

Wilcoxon rank sum test for legumes

```
Alum       N =   4     Median =      2.3700
Clay       N =   4     Median =      2.4550
W = 10.5
Test of ETA1 = ETA2  vs.  ETA1 ~= ETA2 is significant at 0.0433
The test is significant at 0.0421 (adjusted for ties)
```

15.46. Using a Kruskal-Wallis test, we find $H = 9.85$, df $= 2$, and $P = 0.007$. We conclude that there is a difference in iron content for foods cooked in iron pots.

Minitab output: Kruskal-Wallis test for iron pots

```
LEVEL    NOBS    MEDIAN  AVE. RANK   Z VALUE
 Meat       4     4.695       10.5      2.72
  Leg       4     3.705        6.5      0.00
  Veg       4     2.860        2.5     -2.72
OVERALL    12                  6.5

H = 9.85  d.f. = 2  p = 0.007
```

15.47. **(a)** The three pairwise comparisons are *bihai*-red, *bihai*-yellow, and red-yellow. **(b)** The test statistics and P-values are given in the Minitab output below; all P-values are reported as 0 to four decimal places. **(c)** All three are easily significant at the overall 0.05 level.

Minitab output: Wilcoxon rank sum test for *bihai* – red

```
bihai      N =  16     Median =      47.120
red        N =  23     Median =      39.160
W = 504.0
Test of ETA1 = ETA2  vs.  ETA1 ~= ETA2 is significant at 0.0000
```

Wilcoxon rank sum test for *bihai* – yellow

```
bihai      N =  16     Median =      47.120
yellow     N =  15     Median =      36.110
W = 376.0
Test of ETA1 = ETA2  vs.  ETA1 ~= ETA2 is significant at 0.0000
```

Wilcoxon rank sum test for red – yellow

```
red        N =  23     Median =      39.160
yellow     N =  15     Median =      36.110
W = 614.0
Test of ETA1 = ETA2  vs.  ETA1 ~= ETA2 is significant at 0.0000
```

15.48. Multiple comparisons are appropriate as a follow-up to a significant result from a Kruskal-Wallis test, so it only makes sense to do this for Exercises 15.44 and 15.46. That means we have three comparisons from each of these exercises, for a total of 6. In order to be significant at the overall 0.05 level, an individual P-value must be less than $0.05/6 = 0.008\overline{3}$. None of the differences are significant at this level; with such small samples, these tests have low power. (For samples of size 4, W must be between 10 and 26, so five of the six P-values are as small as they can be.)

Minitab output: Medians for the vegetable dishes

```
AlumVeg    N =   4     Median =      1.185
ClayVeg    N =   4     Median =      1.615
IronVeg    N =   4     Median =      2.860
```

Aluminum versus clay pots (vegetable dishes)

```
W = 14.0
Test of ETA1 = ETA2  vs.  ETA1 ~= ETA2 is significant at 0.3123
```

Aluminum versus iron pots (vegetable dishes)

```
W = 10.0
Test of ETA1 = ETA2  vs.  ETA1 ~= ETA2 is significant at 0.0304
```

Clay versus iron pots (vegetable dishes)

```
W = 10.0
Test of ETA1 = ETA2  vs.  ETA1 ~= ETA2 is significant at 0.0304
```

Medians for iron pots

```
IronMeat   N =   4     Median =      4.695
IronLeg    N =   4     Median =      3.705
IronVeg    N =   4     Median =      2.860
```

Meat versus legumes (iron pots)

```
W = 26.0
Test of ETA1 = ETA2  vs.  ETA1 ~= ETA2 is significant at 0.0304
```

Meat versus vegetables (iron pots)

```
W = 26.0
Test of ETA1 = ETA2  vs.  ETA1 ~= ETA2 is significant at 0.0304
```

Legumes versus vegetables (iron pots)

```
W = 26.0
Test of ETA1 = ETA2  vs.  ETA1 ~= ETA2 is significant at 0.0304
```

Chapter 16 Solutions

The solutions for Chapter 16 present a special challenge. Because bootstrap and permutation methods require software, the answers will vary because of (a) random variation due to differences in resampling/rearrangement, and (b) possible systematic and feature differences arising from the specific software used.

Because of (a), most of the solutions here give *ranges* of possible answers, rather than a single answer. These ranges should include the results that most students should get from a single bootstrap or permutation run. (Basically, for each such exercise, I reported the minimum and maximum values from 1000 or more bootstraps or permutations.)

For (b), the text primarily refers to results from S-PLUS, but also mentions SAS and SPSS. If you have other statistical software, you can learn about its bootstrapping capabilities (if any) by consulting your document, or by doing a Web search for the name of your software and "bootstrap." Note that a free student version of S-PLUS is available at `www.onthehub.com/tibco`, so your students may use it for this chapter, even if they normally work with other software. (Faculty can download a 30-day evaluation copy.)

Many of these solutions were originally written by Tim Hesterberg (using S-PLUS) for earlier editions of *IPS,* and have been edited and updated by Darryl Nester using R, the free, open-source version of S-PLUS. One difference in R's bootstrapping library versus that of S-PLUS is that (at the time of this writing) it does not compute "tilting" confidence intervals, so those results are not given. If your software finds tilting intervals, they will (for most of these exercises) be similar to those found by other methods (percentile, BCa, etc.).

16.1. Student answers in this problem will vary substantially due to using different random numbers. (If they do not, you should be suspicious.) **(b)** While students could get a sample mean as low as 0, or as high as 29.78, 95% of all sample means should be between about 5 and 23. **(c)** Shown is a stemplot for a set of 200 resamples. Even for such a large number of resamples, the distribution is somewhat irregular; student stemplots (for 20 resamples) will be even more irregular. **(d)** The theoretical bootstrap standard error is about 4.694, but with only 20 resamples, there will be a fair amount of variation (although almost certainly in the range 2.9 to 6.5).

2	4
3	08
4	
5	09
6	2288
7	011133357
8	0022588
9	13477
10	0002355666889
11	1112233445557777788
12	000000122233334445666688
13	111111224555555688
14	0001144456667799
15	011225588888888889
16	0001444466677
17	133556668
18	02244478
19	0146799
20	001122447788
21	05
22	022225588
23	3
24	9
25	5

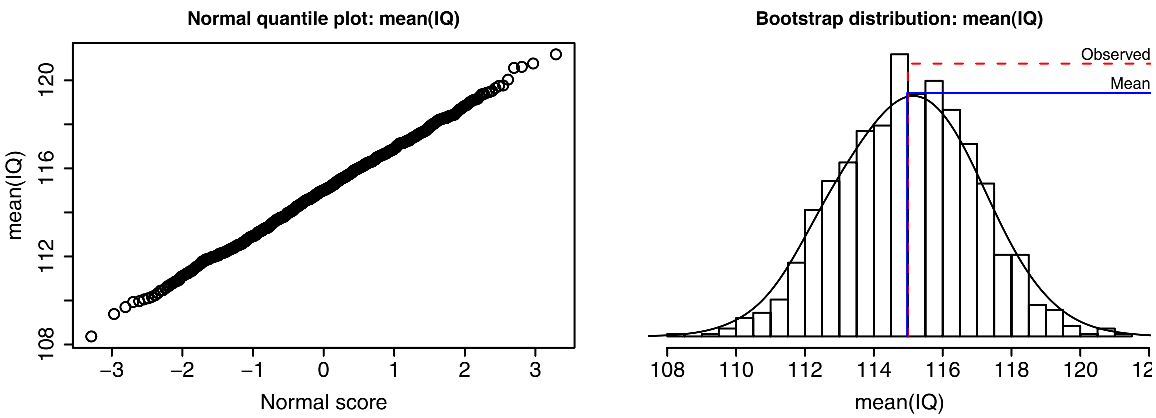
Note: *The range of numbers (5 to 23) given in part (b) is based on 10000 resamples.*

For part (d), the range of standard errors is based on the middle 99% of the SEs from 50000 separate resamples of size 20. The theoretical value is based on considering the six repair times as a population *to compute the standard deviation (dividing by $\sqrt{6}$ rather than $\sqrt{5}$), yielding $\sigma \doteq 11.497$, so the theoretical standard error is $\sigma/\sqrt{6} \doteq 4.694$. The computation in the text (page 16-6) does not mention this detail, although it is discussed briefly in Note 4 on page 16-57. Because bootstrap methods are generally not used with small samples, and the difference is negligible for large samples, it usually does not matter.*

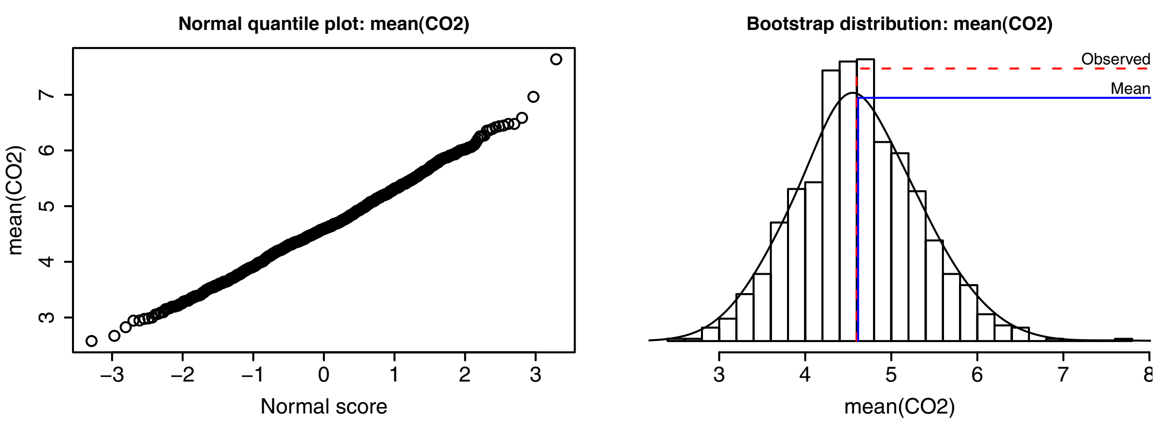
16.2. The standard deviation of a sample measures the spread of that sample. The standard error of a sample mean (or other statistic) estimates how much the mean would vary, if you were to take the means of many samples from the same population. The SE is smaller by a factor of $\sqrt{n}$.

16.3. **(a)** The bootstrap samples from the *sample* (that is, the *data*), not the *population*. **(b)** The bootstrap samples *with* replacement. **(c)** The sample size should be *equal to* the original sample. **(d)** The bootstrap distribution is usually similar to the sampling distribution *in shape and spread, but not in center*.

16.4. The bootstrap distribution is (usually) close to Normal, so we expect the sampling distribution to also be close to Normal.

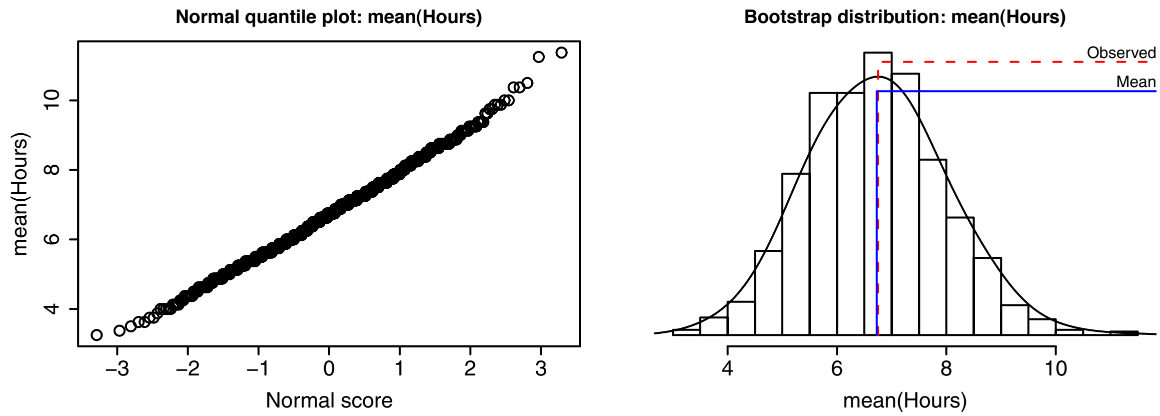


16.5. The bootstrap distribution is (usually) close to Normal, with some positive skewness. We expect the sampling distribution to be close to Normal.

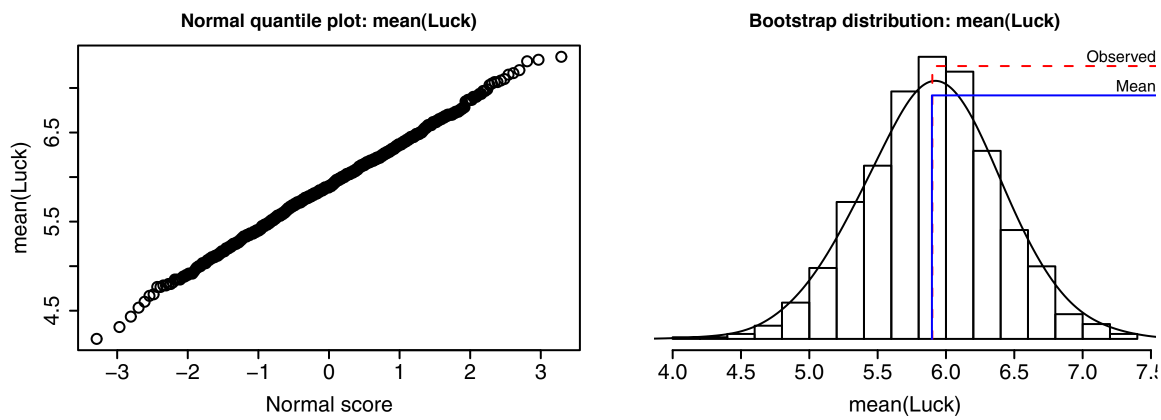


16.6. Due to the small sample size, the bootstrap distribution shows some discreteness (note the small "stair-steps" in the quantile plot). This particular bootstrap distribution looks reasonably Normal, but with a sample size this small, the sample skewness may vary substantially, so we cannot say if the sampling distribution is really skewed.

Note: *The small-sample variability in skewness is not discussed until Section 16.3. Bootstrap methods are not well-suited to datasets this small.*

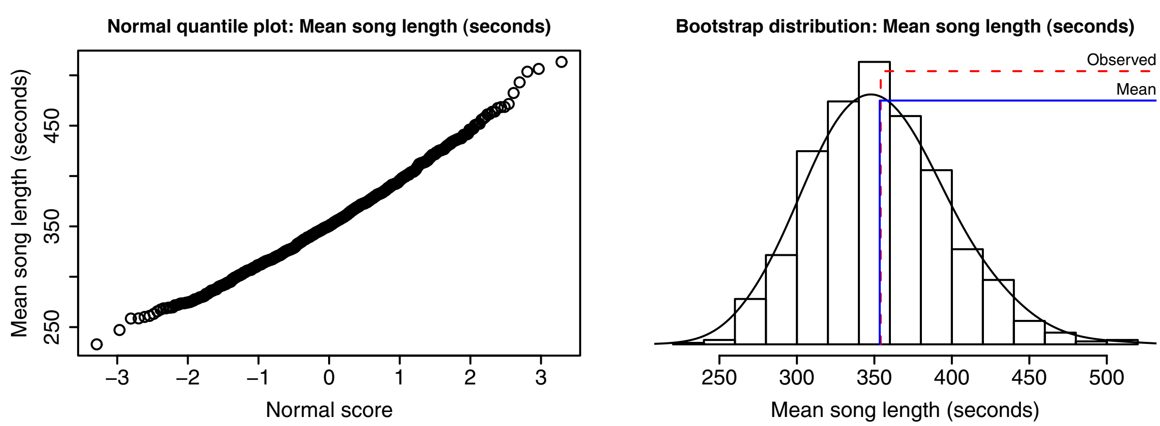


16.7. The bootstrap distribution suggests that the sampling distribution should be close to Normal.



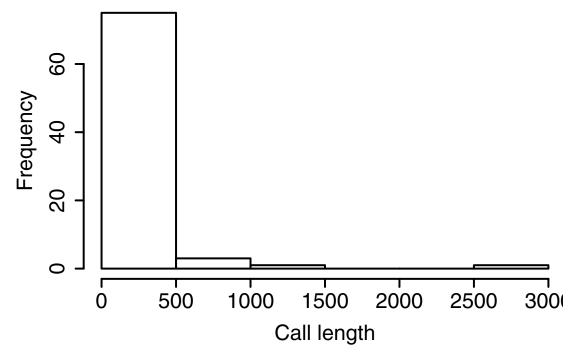
16.8. The bootstrap distribution is non-Normal; specifically, it is skewed to the right. We expect the sampling distribution to be skewed.

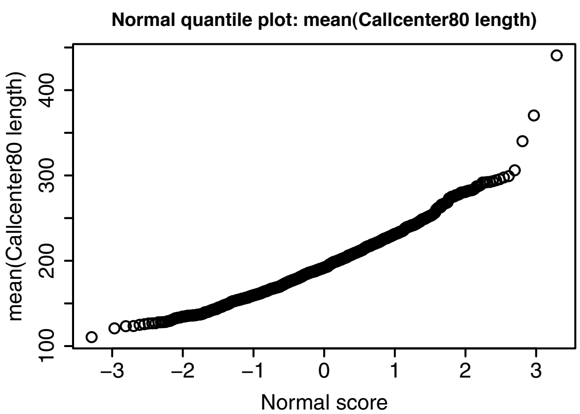
Note: *This amount of skewness would not be a concern with some statistical procedures, but it is when working with bootstrap distributions.*

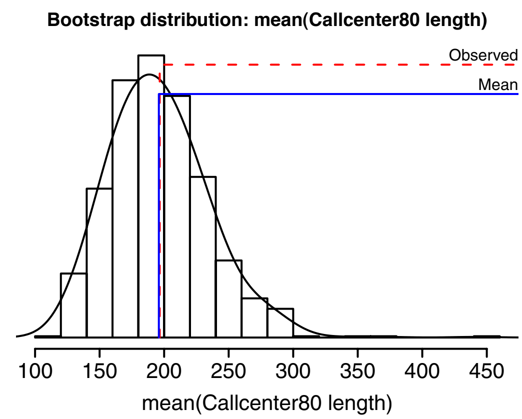


16.9. In each case, SE_{boot} will vary. To get an idea of how much variability one might observe, a range of "typical" bootstrap SE is given, based on 500 trials. **(a)** For the IQ data, $s \doteq 14.8009$, so $SE_{\bar{x}} \doteq 1.9108$. SE_{boot} will typically be between and 1.77 and 2.01. **(b)** For the CO_2 data, $s \doteq 4.8222$, so $SE_{\bar{x}} \doteq 0.6960$. SE_{boot} will typically be between about 0.64 and 0.74. **(c)** For the video-watching data, $s \doteq 3.8822$, so $SE_{\bar{x}} \doteq 1.3726$. SE_{boot} will typically be between about 1.20 and 1.36—almost certainly an underestimate. The bootstrap is biased downward for estimating standard error by a factor of about $\sqrt{(n-1)/n}$, which is about 0.94 when $n = 8$.

16.10. **(a)** Histogram on the right. **(b)** The bootstrap distribution is clearly not Normal in the tails; both the quantile plot and the histogram (on the following page) are clearly skewed to the right.

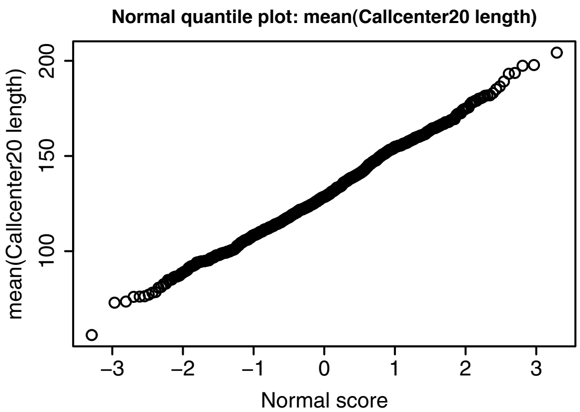


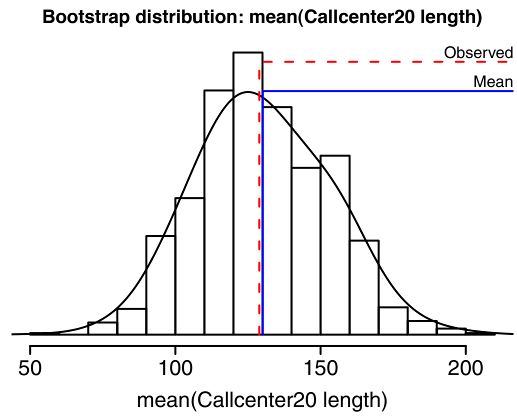




16.11. **(a)** The CALLCENTER20 bootstrap distribution is slightly skewed to the right, but it is considerably *less* skewed than the CALLCENTER80 bootstrap distribution. **(b)** The standard error for the smaller data set is much smaller: For CALLCENTER20, the standard error is almost always between 21 and 25, and for CALLCENTER80, it is almost always between 35 and 41.

Note: *The difference in standard errors is primarily because the sample standard deviation for the CALLCENTER20 data is much smaller (103.8 versus 342.0).*





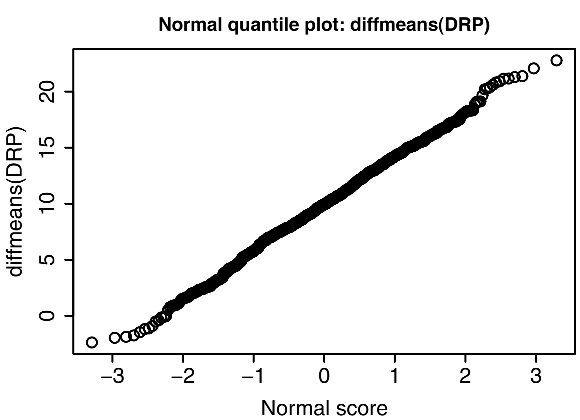
16.12. **(a)** The mean of the sample is $\bar{x} \doteq 13.7567$, so the bootstrap bias estimate is $13.762 - 13.7567 \doteq 0.0053$. **(b)** $\text{SE}_{\text{boot}} = 4.725$. (We do not need to divide the given value by anything; it is already the estimate standard deviation of the sample mean.) **(c)** For df $= 5$, the appropriate critical value is $t^* = 2.571$, so the 95% bootstrap confidence interval is $\bar{x} \pm t^*\text{SE}_{\text{boot}} \doteq 13.7567 \pm 12.146 = 1.6107$ to 25.9027.

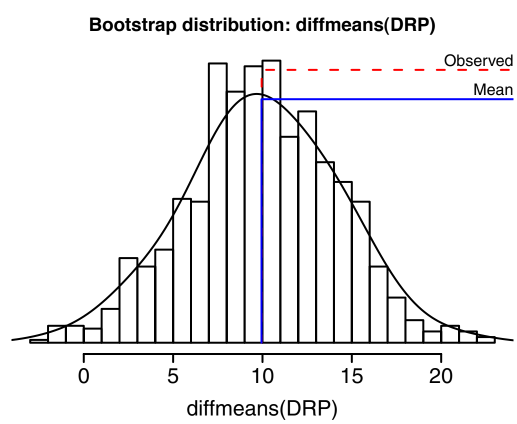
16.13. See also the solution to Exercise 16.8. **(a)** The bootstrap distribution is skewed; a t interval might not be appropriate. **(b)** The bootstrap t interval is $\bar{x} \pm t^*\text{SE}_{\text{boot}}$, where $\bar{x} = 354.1$ sec, $t^* \doteq 2.0096$ for df $= 49$, and SE_{boot} is typically between 39.5 and 46.5. This gives the range of intervals shown on the right. **(c)** The interval reported in Example 7.11 was 266.6 to 441.6 seconds.

Typical ranges	
SE_{boot}	39.5 to 46.5
t lower	260.7 to 274.7
t upper	433.5 to 447.5

16.14. **(a)** Based on 1000 resamples, SE_{boot} is almost always between 3.8 and 4.5. **(b)** The bootstrap distribution looks reasonably Normal, with no appreciable bias, so a bootstrap t interval is appropriate. Typical ranges are on the right. **(c)** The t interval reported in Example 7.15 (page 439) was 1.2 to 18.7.

Typical ranges	
SE_{boot}	3.9 to 4.5
t lower	0.88 to 2.09
t upper	17.82 to 19.03





16.15. The summary statistics given in Example 16.6 include standard deviations $s_1 \doteq 14.7$ min for Verizon and $s_2 \doteq 19.5$ min for CLEC, so $\text{SE}_D \doteq 4.0820$. (Computation from the original data gives $\text{SE}_D \doteq 4.0827$.) The standard error reported by the S-PLUS bootstrap routine (shown in the text following that example) is 4.052.

16.16. See also the solution to Exercise 16.6. **(a)** The bootstrap distribution found in the solution to Exercise 16.6 looks reasonably Normal. **(b)** The likely range of bootstrap intervals is on the right. **(c)** With $\bar{x} = 6.75$ hr, $s \doteq 3.8822$, and $t^* \doteq 2.365$ (df = 7), the usual 95% confidence interval is 3.50 to 10.00, so the bootstrap interval is almost always narrower than the usual interval.

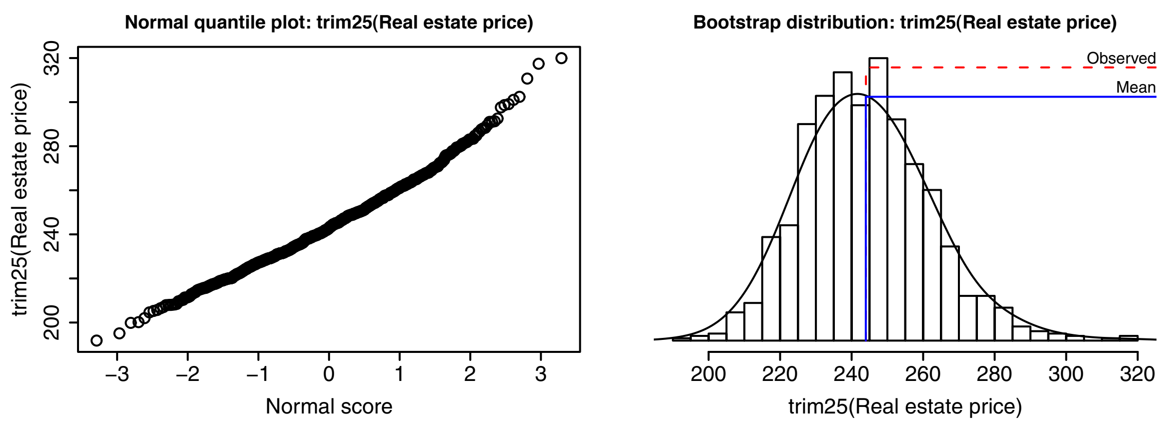
Typical ranges	
SE_{boot}	1.2 to 1.4
t lower	3.5 to 3.9
t upper	9.6 to 10

16.17. See also the solution to Exercise 16.10. **(a)** The bootstrap bias is typically between −4 and 4, which is small relative to $\bar{x} = 196.575$ min. **(b)** Ranges for the bootstrap interval are given on the right. **(c)** $\text{SE}_{\bar{x}} \doteq 38.2392$, while SE_{boot} ranges from about 35 to 41. The usual t interval is 120.46 to 272.69 min.

Typical ranges	
Bias	−4 to 4
SE_{boot}	35 to 41
t lower	114 to 127
t upper	266 to 279

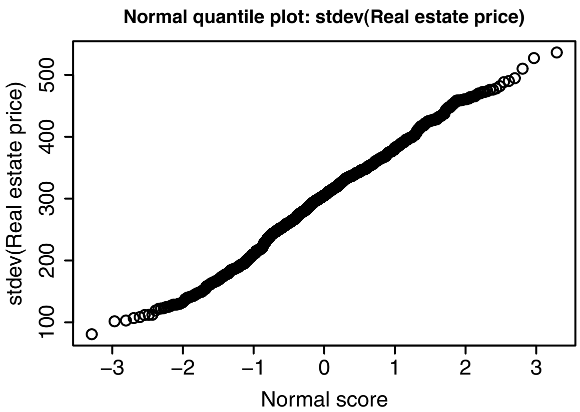
16.18. **(a)** The two distributions should be similar (although the default plots may differ depending on what software is used). **(b)** Typical ranges for these quantities are shown on the right. **(c)** The interval in Example 16.5 was (210.19, 277.81).

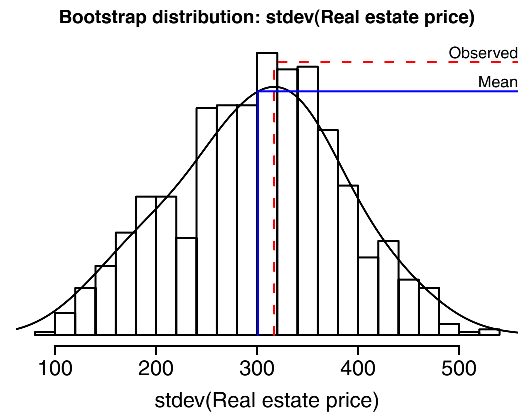
Typical ranges	
$\bar{x}_{25\%}$	242.7 to 246.4
Bias	−1.3 to 2.4
SE_{boot}	16 to 19
t lower	205.5 to 211.5
t upper	276.5 to 282.5



16.19. The bootstrap distribution (below) is noticeably non-Normal in the tails, especially the low tail. In addition, the bias is larger than we would like (and almost always negative, because of the heavy low tail). A t interval is risky (and perhaps not appropriate), but students may elect to compute it anyway.

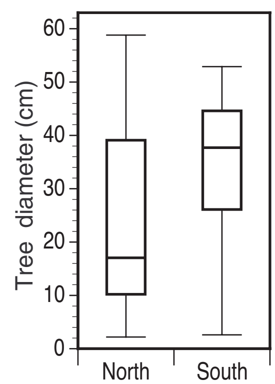
Typical ranges	
Bias	−22.6 to −7.7
SE_{boot}	76.7 to 87.9
t lower	140.3 to 162.6
t upper	471.1 to 493.4



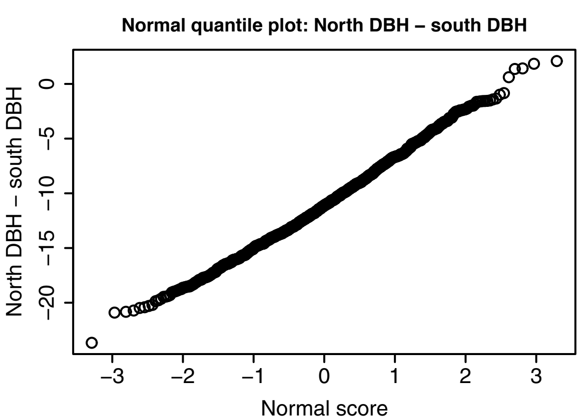


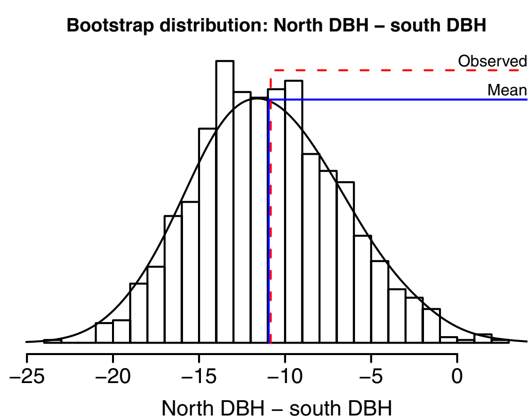
16.20. **(a)** Shown on the right are both stemplots and boxplots (copied from the solution to Exercise 7.81). The north distribution is right-skewed, while the south distribution is left-skewed. It might be appropriate to use standard t methods in spite of the skewness because the sample sizes are relatively large, and there are no outliers in either distribution. **(b)** The bootstrap distribution appears to be quite close to Normal, with very little bias. **(c)** Typical ranges for SE_{boot} and the bootstrap interval are given on the right. **(d)** The standard t interval is -19.09 to -2.58 (df $= 55.7$).

North		South
43322	0	2
65	0	57
443310	1	2
955	1	8
	2	13
8755	2	689
0	3	2
996	3	566789
43	4	003444
6	4	578
4	5	0112
85	5	



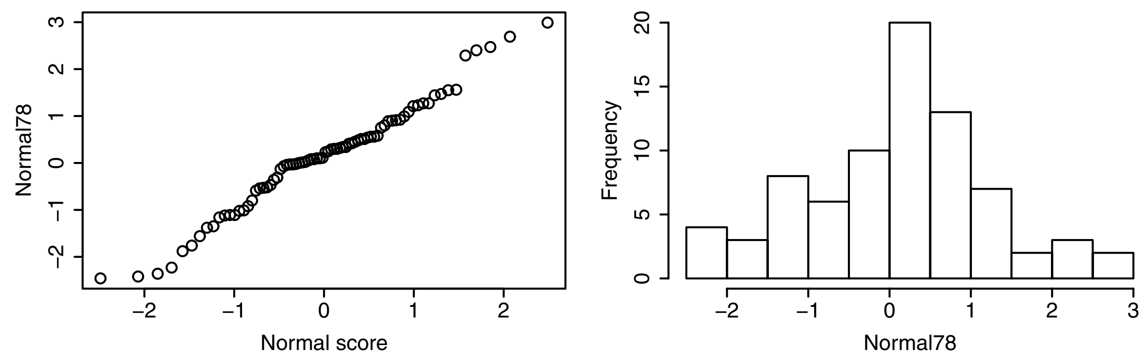
Typical ranges	
Bias	-0.4 to 0.5
SE_{boot}	3.7 to 4.3
t lower	-19.5 to -18.3
t upper	-3.4 to -2.2

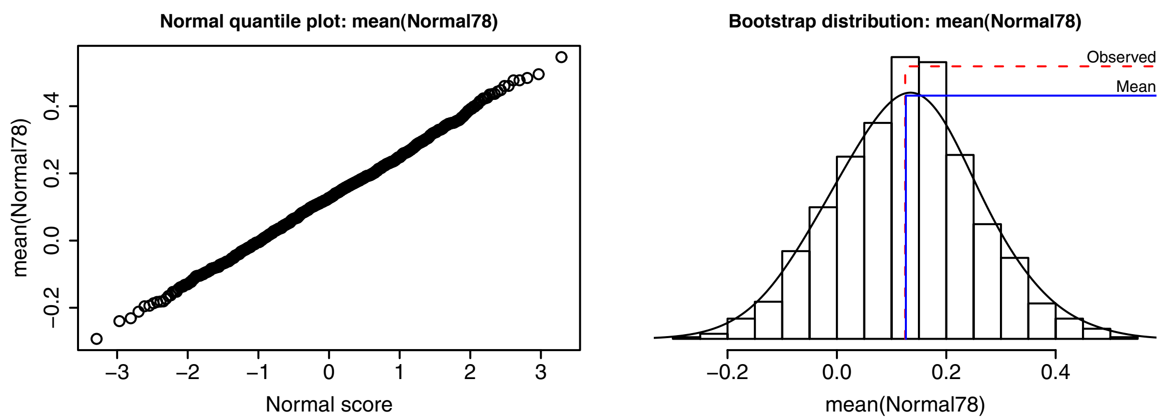




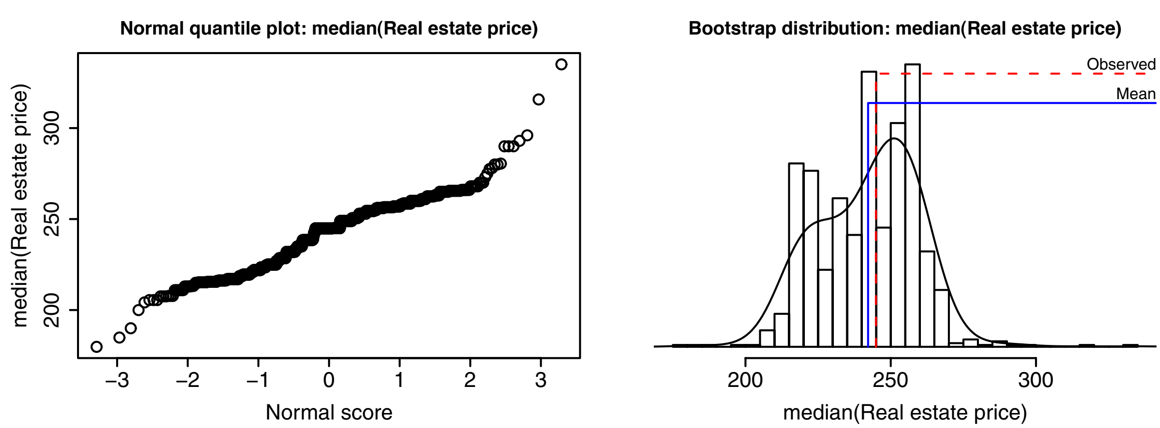
16.21. **(a)** The data appear to be roughly Normal, though with the typical random gaps and bunches that usually occur with relatively small samples. It appears from both the histogram and quantile plot that the mean is slightly larger than zero, but the difference is not large enough to rule out a $N(0, 1)$ distribution. **(b)** The bootstrap distribution is extremely close to Normal with no appreciable bias. **(c)** $\text{SE}_{\bar{x}} \doteq 0.1308$, and the usual t interval is -0.1357 to 0.3854. Typical results for SE_{boot} and the bootstrap interval are above on the right.

Typical ranges	
Bias	-0.016 to 0.02
SE_{boot}	0.12 to 0.14
t lower	-0.16 to -0.11
t upper	0.36 to 0.41



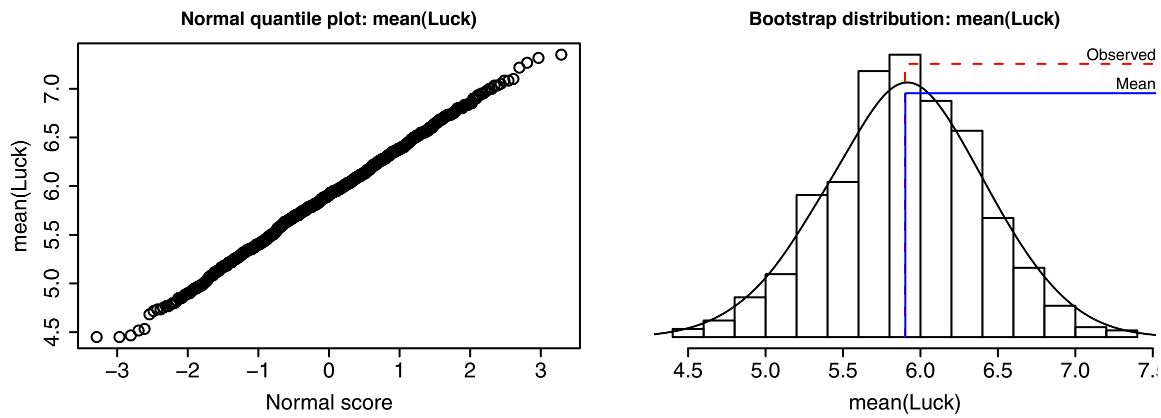


16.22. Based on a quantile plot and histogram, the bootstrap distribution is quite non-Normal.



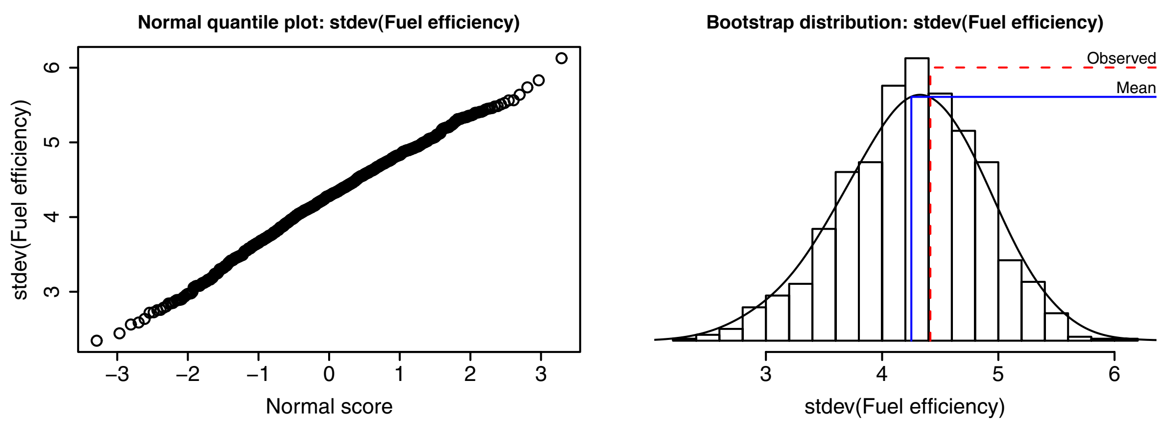
16.23. Because the scores are all between 1 and 10, there can be no extreme outliers, so standard t methods should be safe. The bootstrap distribution appears to be quite Normal, with little bias. The usual t interval is 4.9256 to 6.8744, which is in the range of typical bootstrap intervals.

Typical ranges	
Bias	−0.07 to 0.06
SE_{boot}	0.44 to 0.53
t lower	4.8 to 5.0
t upper	6.8 to 7.0



16.24. **(a)** The sample standard deviation is $s \doteq 4.4149$ mpg. **(b)** The typical range for SE_{boot} is in the table on the right. **(c)** SE_{boot} is quite large relative to s, suggesting that s is not a very accurate estimate. **(d)** There is substantial negative bias and some skewness, so a t interval is probably not appropriate.

Typical ranges	
Bias	−0.22 to -0.09
SE_{boot}	0.55 to 0.65
t lower	3.07 to 3.26
t upper	5.57 to 5.76

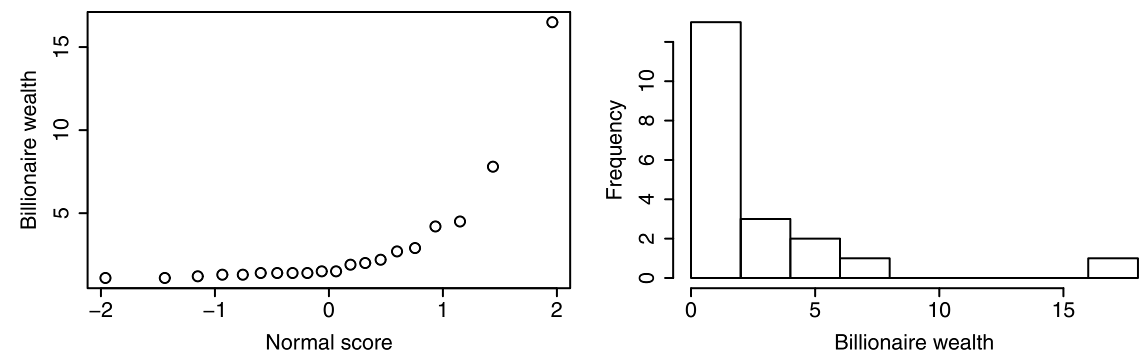


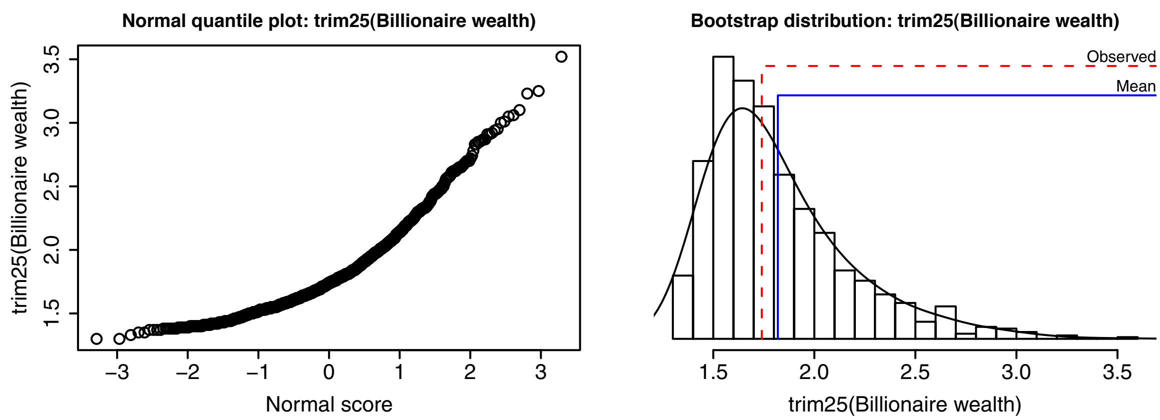
16.25. The distribution is sharply right-skewed, so a trimmed mean is a good choice. (Students will likely choose the 25% trimmed mean, both because it is mentioned in the text, and because it is given in the answer to this exercise.) For the sample, $\bar{x}_{25\%} = 1.74$ billion dollars. Typical ranges for the bias and SE_{boot} are given on the right; note that the bias is a substantial fraction of SE_{boot}.

Typical ranges	
Bias	0.048 to 0.114
SE_{boot}	0.29 to 0.41
t lower	0.89 to 1.12
t upper	2.36 to 2.59

The bootstrap distribution is strongly right-skewed, so a bootstrap t interval would be questionable; the typical range of intervals is shown (above, right).

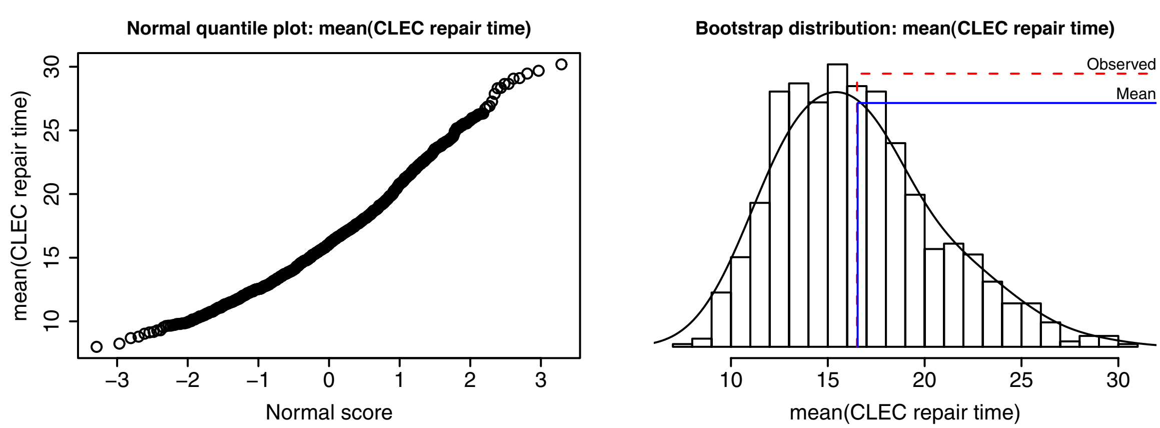
Note: *By definition, the mean wealth (trimmed or not) of billionaires must be more than \$1 billion; the fact that the bootstrap interval can extend below that limit is an indication that we should not rely on it.*



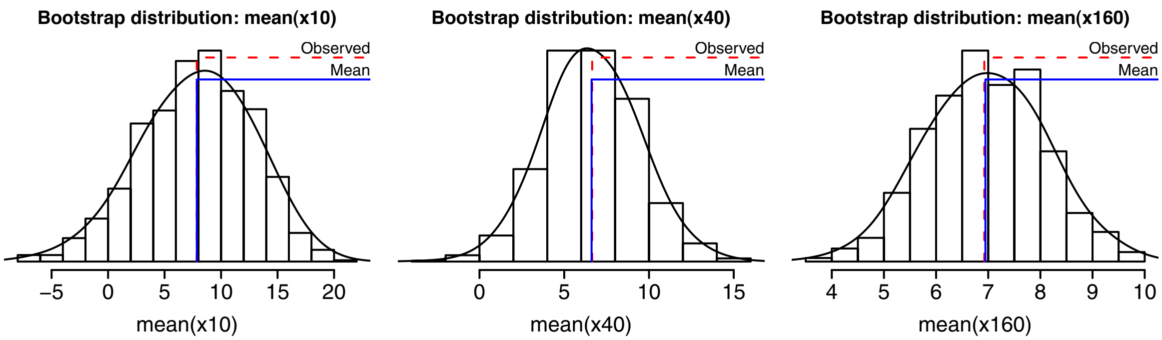


16.26. **(a)** The bootstrap distribution for the CLEC mean is strongly right-skewed with mean 16.5, with bias and $\mathrm{SE_{boot}}$ in the ranges shown on the right. For comparison, the bootstrap distribution for ILEC mean in Figure 16.3 is barely skewed, with a mean of 8.41 and $\mathrm{SE_{boot}} = 0.367$. **(b)** Note that $\mathrm{SE_{boot}}$ is much larger for CLEC than for ILEC. Because the ILEC bootstrap means vary so little, when we compute $\bar{x}_{\mathrm{ILEC}} - \bar{x}_{\mathrm{CLEC}}$, it is the latter term that primarily determines the shape of the distribution of the difference. Because of the minus sign, the right skewness of the CLEC means cause the difference to be left-skewed.

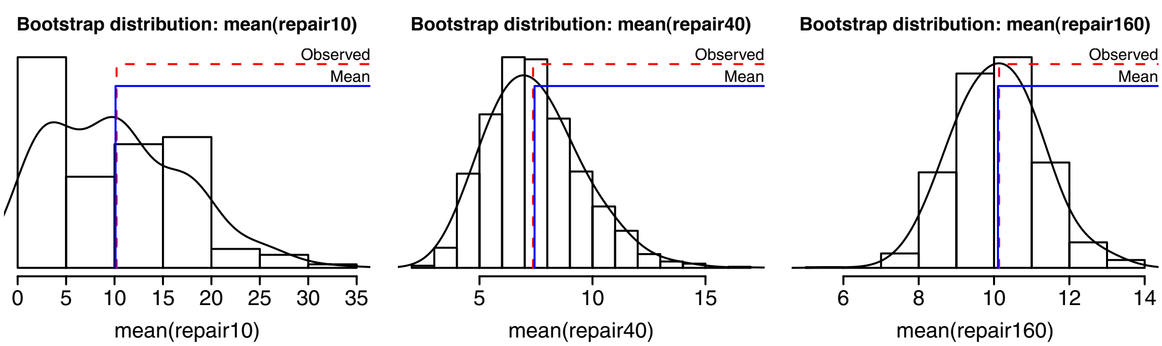
Typical ranges	
Bias	−0.40 to 0.42
$\mathrm{SE_{boot}}$	3.6 to 4.3
t lower	7.6 to 9.0
t upper	24.0 to 25.4



16.27. **(a)** $\bar{x}$ would have a Normal distribution with mean 8.4 and standard deviation $14.7/\sqrt{n}$. **(b)** and **(c)** Histograms below. The values of SE_{boot} will be quite variable, both because of variation in the original sample, and variation due to resampling. **(d)** Student answers will vary, depending on their samples. There may be some skewness (right or left) for smaller samples. SE_{boot} should be roughly halved each time the sample size increases by a factor of 4, although for $n = 10$ and $n = 40$, the size of SE_{boot} can vary considerably.



16.28. **(a)** The mean is 8.4, and the standard deviation is $14.7/\sqrt{n}$. **(b)** and **(c)** Histograms on the following page. The values of SE_{boot} will be quite variable, both because of variation in the original sample, and variation due to resampling. **(d)** Student answers may vary, depending on their samples. There may be substantial right skewness, and some irregularity, for smaller samples (the $n = 10$ histogram below is an extreme case), but the distribution should be closer to Normal for large samples. There will typically be little or no bias. SE_{boot} should be roughly halved each time the sample size increases by a factor of 4, although for $n = 10$ and $n = 40$, the size of SE_{boot} may vary considerably.



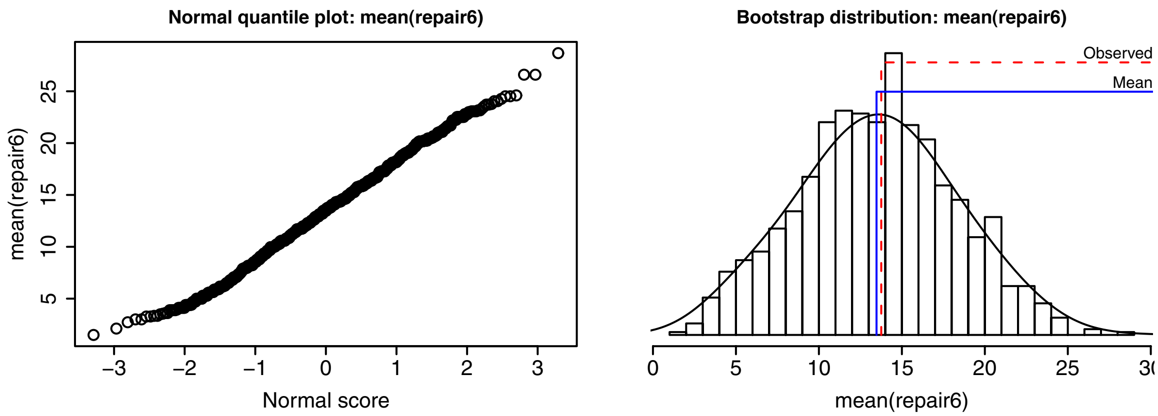
16.29. Student answers should vary depending on their samples. They should notice that the bootstrap distributions are approximately Normal for larger sample sizes. For small samples, the sample could be skewed one way or the other in Exercise 16.27, and most should be right skewed for Exercise 16.28. Some of that skewness should come through into the bootstrap distribution.

16.30. For a 90% bootstrap percentile confidence interval, we choose the 5th and 95th percentiles of the bootstrap distribution. For an 80% interval, use the 10th and 90th percentiles.

16.31. **(a)** The bootstrap distribution looks close to Normal (though that does not mean much with this small sample). The bias is small. **(b)** The typical range of bootstrap t intervals is on the right. **(c)** The bootstrap percentile interval is much narrower than the bootstrap t interval. (It is typically 70% to 80% as wide.)

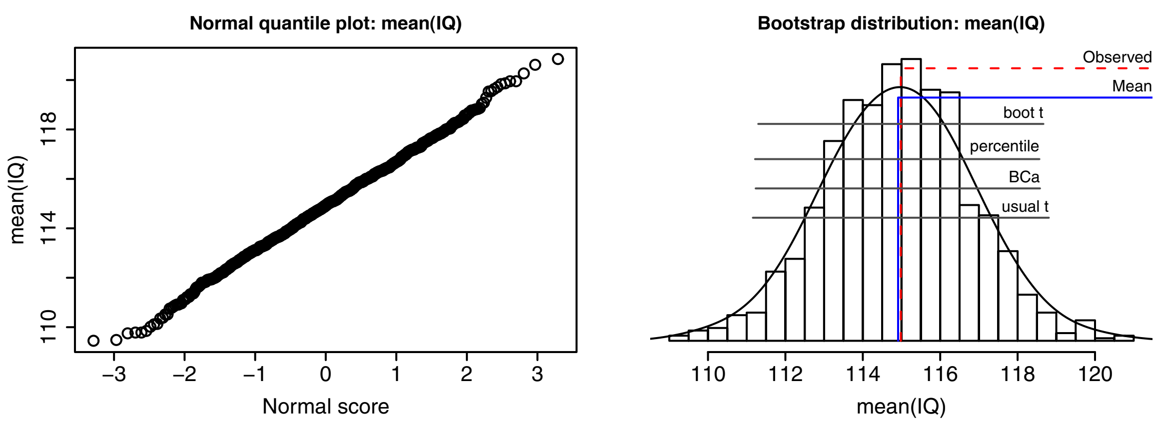
Typical ranges	
Bias	−0.54 to 0.54
SE_{boot}	4.35 to 5.01
t lower	0.89 to 2.57
t upper	24.94 to 26.63
Percentile lower	3.7 to 5.9
Percentile upper	22.1 to 24.5

Note: *The reason that the percentile interval is narrower in this setting is that the bootstrap distribution has "heavy tails," visible in the slight curvature on the edges of the quantile plot. This inflates the standard error, and therefore make the t interval wider than it should be.*



16.32. **(a)** The sample standard deviation is $s \doteq 14.8$ with $n = 60$, so $SE_{\bar{x}} \doteq 1.9108$. The usual t confidence interval is $\bar{x} \pm 2.001\, SE_{\bar{x}} \doteq 114.9833 \pm 3.8235 = 111.16$ to 118.81. **(b)** The bootstrap distribution appears to be reasonably close to Normal, except perhaps in the tails. Ranges for SE_{boot} and the bootstrap t interval are given on the right. **(c)** The intervals agree fairly well, although resampling variation can produce different results. Here the formula interval would be fine.

Typical ranges	
Bias	−0.19 to 0.23
SE_{boot}	1.78 to 2.04
t lower	110.9 to 111.5
t upper	118.5 to 119.1



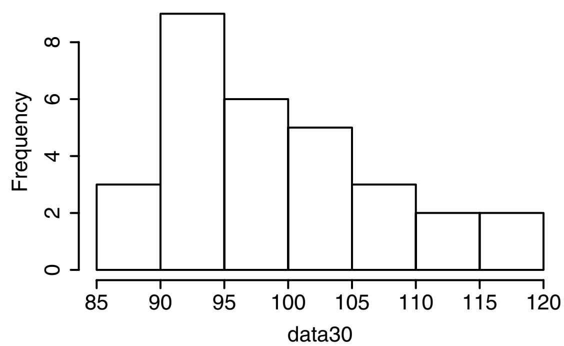
16.33. (a) The bootstrap percentile and t intervals are very similar, suggesting that the t intervals are acceptable. **(b)** Every interval (percentile and t) includes 0.

Typical ranges	
t lower	−0.16 to −0.11
t upper	0.36 to 0.41
Percentile lower	−0.17 to −0.09
Percentile upper	0.35 to 0.42

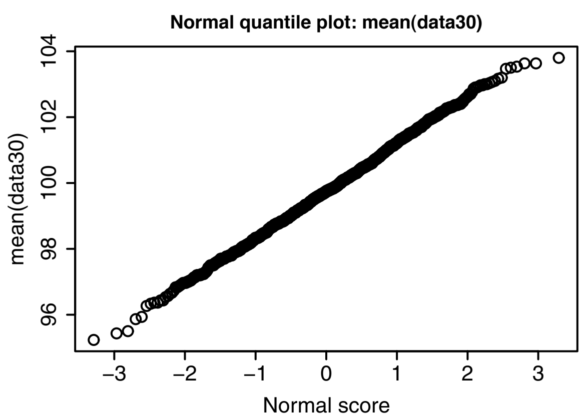
Note: *In the solution to Exercise 16.31, the percentile intervals were always 70% to 80% as wide as the t intervals (because of the heavy tails of that bootstrap distribution). In this case, the width of the percentile interval is 93% to 106% of the width of the t interval.*

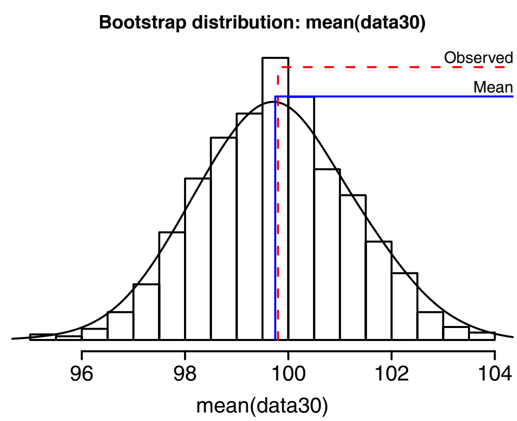
16.34. (a) The distribution is skewed to the right, but has no extreme outliers, so standard t procedures should be safe unless high accuracy is needed. **(b)** The usual t interval is $\bar{x} \pm 2.045\text{SE}_{\bar{x}} = 99.8 \pm 3.05 = 96.75$ to 102.85. **(c)** The bootstrap distribution is very close to Normal with no appreciable bias; a t interval should be accurate. Ranges are in the table on the right. **(d)** The bootstrap percentile interval (ranges on the right) is typically similar to the bootstrap t interval, and both are similar to the standard t interval. We conclude that the usual t interval is accurate.

Note: *The width of the percentile interval is typically 90% to 103% of the width of the t interval.*



Typical ranges	
Bias	−0.15 to 0.15
SE_{boot}	1.34 to 1.60
t lower	96.7 to 97.0
t upper	102.6 to 102.9
Percentile lower	96.6 to 97.4
Percentile upper	102.3 to 103.2





16.35. These intervals are given on page 16-37 of the text: The percentile interval is (−0.128, 0.356), and the bootstrap t interval is (−0.144, 0.358). The differences are relatively small relative to the width of the intervals, so they do not indicate appreciable skewness.

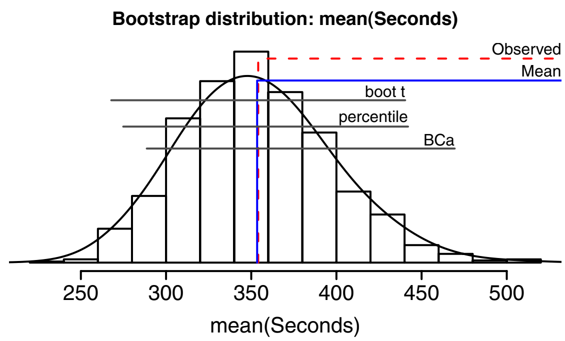
16.36. **(a)** These intervals will vary in a manner similar to the bootstrap t and percentile intervals; see the solution to Exercise 16.34 for likely ranges. **(b)** BCa and tilting intervals are typically similar to the standard t interval (96.75 to 102.85), again suggesting that the usual t interval is accurate. **(c)** For a quick check, we might use the percentile interval. For a more accurate check, we should use a BCa or tilting interval (if they are available).

16.37. Typical ranges for the BCa interval are shown on the right; the tilting interval will be similar. Most intervals are fairly similar to the bootstrap t and percentile intervals from Example 16.10, suggesting that the simpler intervals are adequate.

Typical ranges	
BCa lower	-0.19 to -0.11
BCa upper	0.32 to 0.41

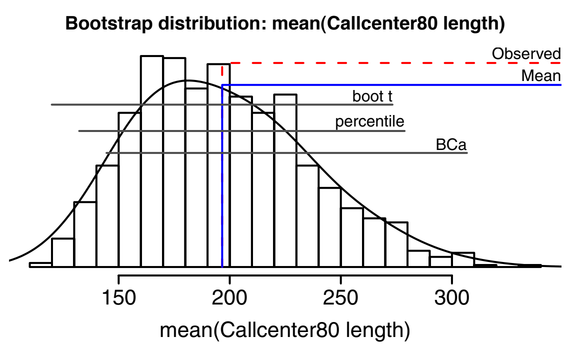
16.38. As was previously noted in the solutions to Exercises 16.8 and 16.13, the skewness is a cause for concern. The lower limit of the percentile interval is generally larger than lower limit of the bootstrap t interval. The BCa interval is almost always shifted to the right relative to both the t and percentile intervals. The t and percentile intervals are inaccurate here; the more sophisticated BCa or tilting intervals are more reliable.

Typical ranges	
t lower	260.7 to 274.7
t upper	433.5 to 447.5
Percentile lower	270.2 to 287.0
Percentile upper	432.8 to 463.3
BCa lower	281.4 to 300.0
BCa upper	450.9 to 516.2



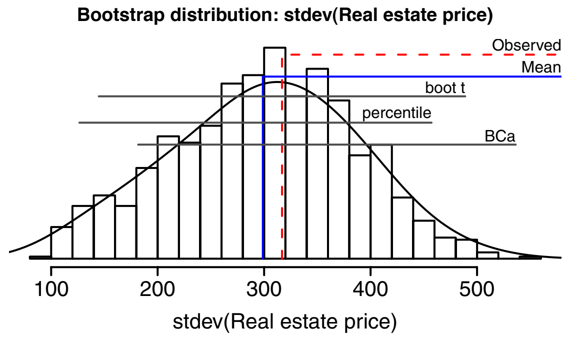
16.39. The percentile interval is shifted to the right relative to the bootstrap t interval. The more accurate intervals are shifted even further to the right.

Typical ranges	
t lower	114 to 127
t upper	266 to 279
Percentile lower	127 to 140
Percentile upper	267 to 298
BCa lower	137 to 152
BCa upper	292 to 371



16.40. The percentile interval is typically shifted to the left relative to the others; the BCa and tilting intervals are farther to the right. Based on these differences, the BCa or tilting interval should be used.

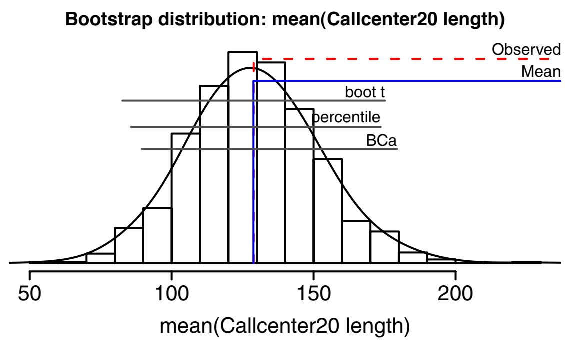
Typical ranges	
t lower	139.4 to 162.7
t upper	470.9 to 494.2
Percentile lower	122.8 to 151.3
Percentile upper	437.7 to 479.5
BCa lower	162.3 to 202.3
BCa upper	484.7 to 609.1



16.41. The bootstrap distribution for the smaller sample is *less* skewed. The standard t interval is 80.34 to 177.46; the bootstrap t interval is similar, and the other bootstrap intervals are generally narrower and shifted to the right.

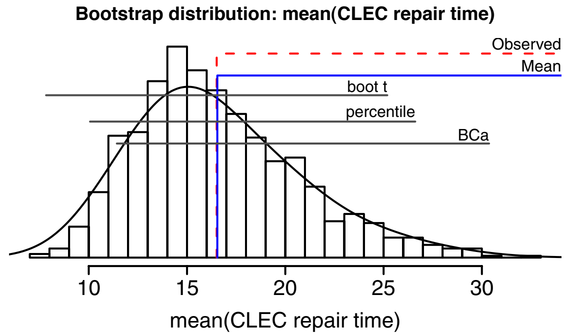
Note: *Generally, a smaller sample should result in* less *regularity—that is, more skewness, larger standard error, etc. In this case, the smaller sample does not contain the nine highest call lengths, many of which would be considered outliers. Those increase the skewness of the bootstrap distribution for CALLCENTER80.*

Typical ranges	
t lower	78.0 to 84.4
t upper	173.3 to 179.8
Percentile lower	80.8 to 92.8
Percentile upper	169.7 to 181.6
BCa lower	83.1 to 96.2
BCa upper	173.0 to 191.9



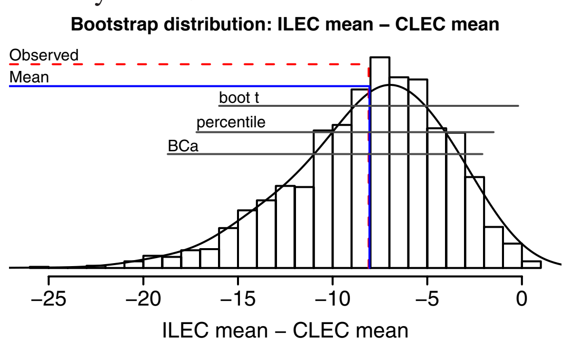
16.42. See also the solution to Exercise 16.26. **(a)** The bootstrap distribution shows strong right skewness, making the formula t, bootstrap t, and (to a lesser extent) the percentile intervals unreliable. The BCa and tilting intervals adjust for skewness, so they should be accurate. **(b)** Typical ranges for these intervals are shown on the right. Relative to the sample mean $\bar{x} \doteq 16.5$, the BCa and tilting intervals are much more asymmetrical than the other intervals, because they take into account the skewness in the data. The bootstrap t ignores the skewness, and the percentile interval only catches part of the skewness. In practical terms, a t interval would tend to underestimate the true value: It would not stretch far enough to the right, so it would have a probability of missing the population mean higher than 5%. This is true to a lesser extent for the percentile interval.

Typical ranges	
t lower	7.5 to 9.00
t upper	24.0 to 25.5
Percentile lower	9.5 to 10.6
Percentile upper	23.9 to 27.3
BCa lower	10.7 to 12.0
BCa upper	26.8 to 37.9



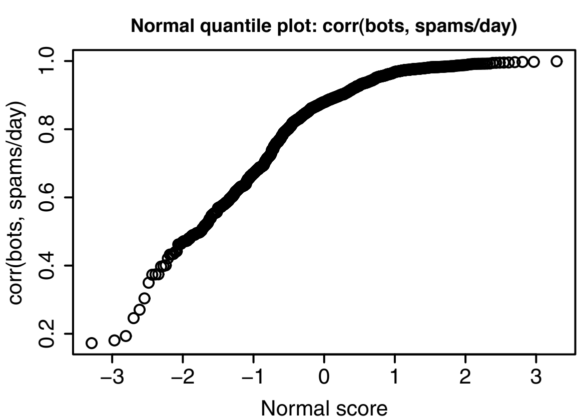
16.43. The observed difference is $\bar{x}_{\text{ILEC}} - \bar{x}_{\text{CLEC}} \doteq -8.1$. Ranges for all three intervals are given below. Because of the left skew of the bootstrap distribution, the t interval does not reach far enough to the left and reaches too far to the right, meaning that the interval would be too high too often, effectively overestimating where the true difference lies. This may also be true for the percentile interval, but considerably less so.

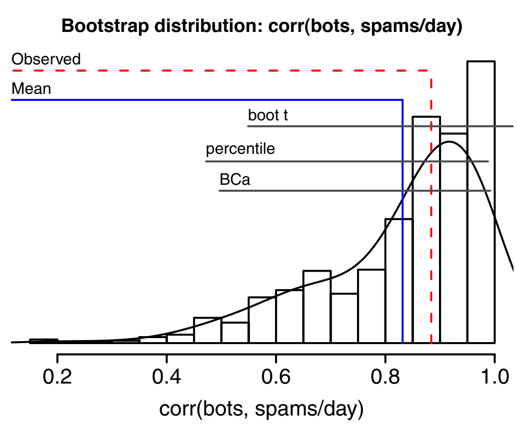
Typical ranges	
t lower	-16.5 to -15.4
t upper	-0.8 to 0.3
Percentile lower	-18.5 to -16.0
Percentile upper	-2.1 to -1.2
BCa lower	-19.3 to -16.5
BCa upper	-2.5 to -1.4



16.44. (a) The bootstrap distribution is *extremely* left-skewed, with consistent negative bias. This is definitely not a candidate for the bootstrap t interval. **(b)** As expected, the t interval is not acceptable; in particular, the upper limit is always greater than 1. (In the plot below, it extends past the right border of the histogram.) The percentile and BCa intervals are often similar; while the BCa interval is theoretically more sophisticated, in this case, the simpler percentile method seems to be fine.

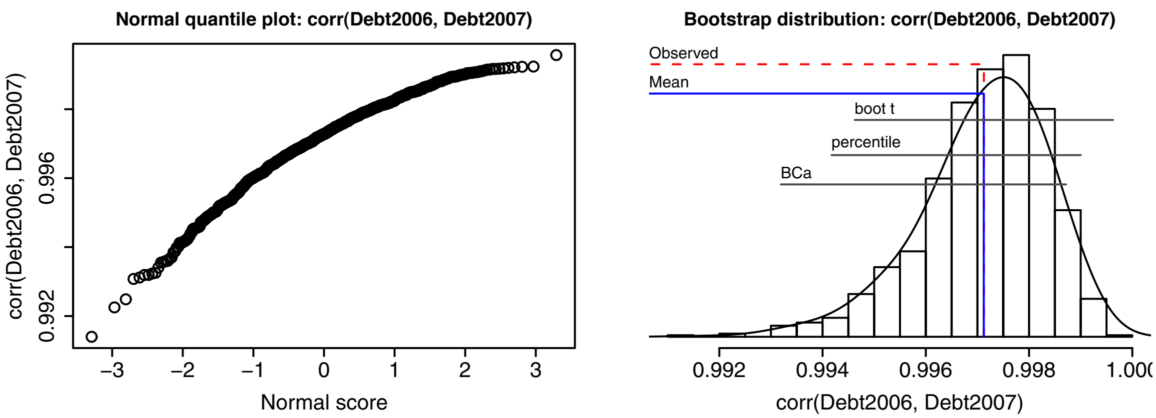
Typical ranges	
Bias	-0.071 to -0.033
SE_{boot}	0.13 to 0.19
t lower	0.47 to 0.57
t upper	1.19 to 1.30
Percentile lower	0.37 to 0.51
Percentile upper	0.98 to 1.00
BCa lower	0.38 to 0.55
BCa upper	0.98 to 1.00





16.45. **(a)** The bootstrap distribution is sharply left-skewed. **(b)** Shown are ranges for the percentile and BCa intervals, as well as the (inappropriate) bootstrap t interval. The percentile and BCa intervals typically have similar upper limits, but the BCa lower limit is generally less than the percentile lower limit. The confidence intervals give more than enough evidence to reject H_0: $\rho = 0$; in fact, we have strong evidence that the correlation is at least 0.99.

Typical ranges	
Bias	−0.00019 to 0.00008
t lower	0.9940 to 0.9948
t upper	0.9995 to 1.0002
Percentile lower	0.9931 to 0.9946
Percentile upper	0.9988 to 0.9992
BCa lower	0.9899 to 0.9937
BCa upper	0.9985 to 0.9990



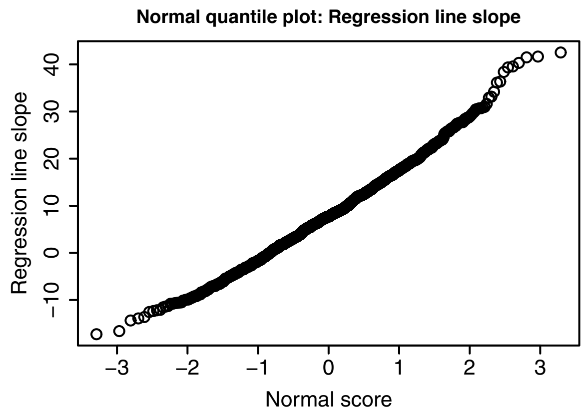
16.46. We should resample whole observations. If the data are stored in a spreadsheet with observations in rows and the x and y variables in two columns, then we should pick a random sample of rows with replacement. When a row is picked, we put the whole row into a bootstrap data set. By doing so, we maintain the relationship between x and y.

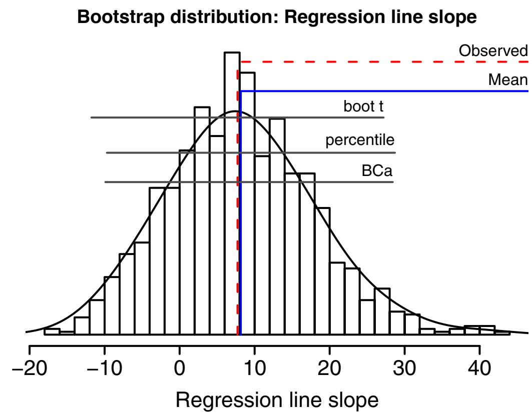
16.47. **(a)** The regression equation for predicting salary (in \$millions) is $\hat{y} = 0.8125 + 7.7170x$, where x is batting average. (The slope is *not* significantly different from 0: $t = 0.744$, $P = 0.461$.) **(b)** The bootstrap distribution is reasonably close to Normal, suggesting that any of the intervals would be reasonably accurate. Ranges for the intervals are given on the right. **(c)** These results are consistent with our conclusion about correlation: All intervals include zero, which corresponds to no (linear) relationship between batting average and salary.

Typical ranges	
Bias	−0.56 to 1.35
SE_{boot}	9.26 to 10.6
t lower	−13.6 to −10.8
t upper	26.3 to 29.0
Percentile lower	−12.7 to −7.7
Percentile upper	26.0 to 31.7
BCa lower	−13.6 to −7.6
BCa upper	24.7 to 33.0

R output: `lm(I(Salary/10^6) ~ Average, data=mlbsalaries)`

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.8125     2.6941   0.302    0.764
Average       7.7170    10.3738   0.744    0.461

Residual standard error: 2.718 on 48 degrees of freedom
Multiple R-squared: 0.0114,     Adjusted R-squared: -0.009199
F-statistic: 0.5534 on 1 and 48 DF,  p-value: 0.4606
```





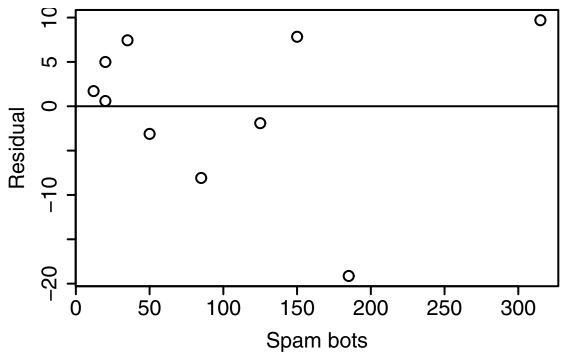
16.48. **(a)** The residuals versus bots plot suggests that spread increases with the number of bots—a violation of the conditions for regression. The small sample size makes it hard to draw conclusions from the quantile plot (it is not very linear, but that is primarily because of one point). **(b)** The bootstrap distribution is decidedly non-Normal; we should not use the t interval. (The percentile interval is more reliable, but is still somewhat risky because of the shape of the bootstrap distribution.) **(c)** Ranges for all three intervals are given in the table on the right. As expected, the bootstrap t interval is inaccurate, but the percentile and BCa intervals are quite similar. With $t^* = 2.306$ for df $= 8$, the standard t interval is $0.1705 \pm t^*(0.03189) = 0.0969$ to 0.2440—shifted to the right relative to BCa (and percentile).

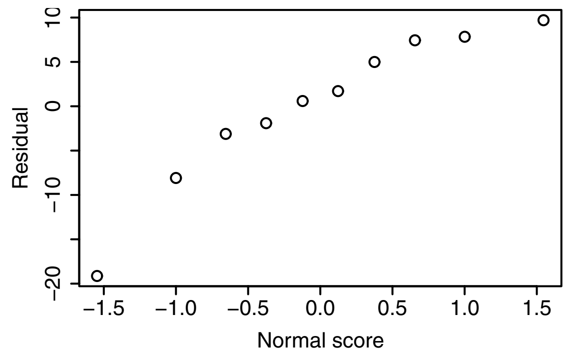
Typical ranges	
Bias	−0.02 to −0.01
SE_{boot}	0.045 to 0.053
t lower	0.050 to 0.067
t upper	0.274 to 0.290
Percentile lower	0.028 to 0.049
Percentile upper	0.207 to 0.216
BCa lower	0.029 to 0.053
BCa upper	0.207 to 0.224

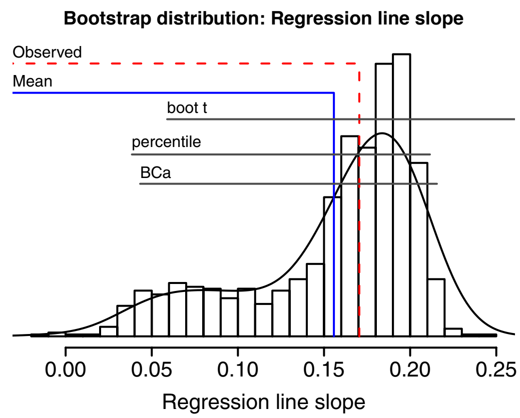
Note: *Because H_0: $\beta_1 = 0$ is equivalent to H_0: $\rho = 0$, it might be tempting to think that bootstrapping the slope and bootstrapping the correlation should be equivalent. Although the distributions are usually similar, the resampling process also changes s_x and s_y, which complicates the relationship between ρ and β_1.*

R output: `lm(formula = SpamsPerDay ~ Bots, data = spam)`

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -3.40192    4.31939  -0.788 0.453635
Bots         0.17048    0.03189   5.346 0.000689

Residual standard error: 9.246 on 8 degrees of freedom
Multiple R-squared: 0.7813,     Adjusted R-squared: 0.754
F-statistic: 28.58 on 1 and 8 DF,  p-value: 0.0006891
```





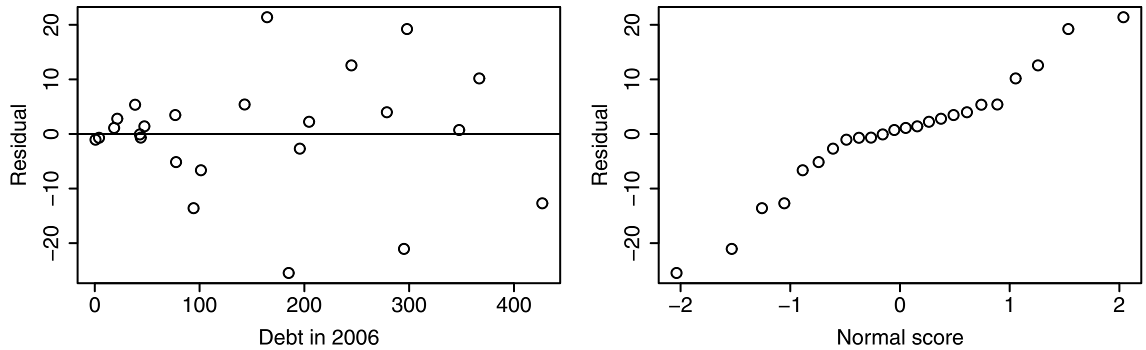


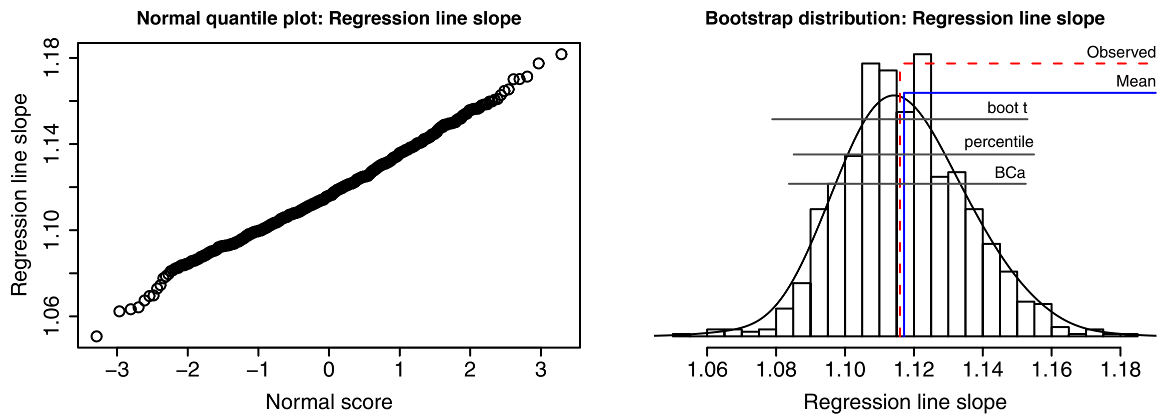
16.49. **(a)** The tails of the residuals are somewhat heavier than we would expect from a Normal distribution. **(b)** The bootstrap distribution is reasonable close to Normal, so the bootstrap t should be fairly accurate. **(c)** Ranges for all three bootstrap intervals are given in the table on the right; they all give similar results. With $t^* = 2.074$ for df $= 22$, the standard t interval is $1.1159 \pm t^*(0.0108) = 1.0784$ to 1.1534. The bootstrap intervals are fairly close to this.

Typical ranges	
Bias	−0.0008 to 0.0026
SE_{boot}	0.016 to 0.019
t lower	1.07 to 1.09
t upper	1.14 to 1.16
Percentile lower	1.08 to 1.09
Percentile upper	1.14 to 1.16
BCa lower	1.07 to 1.09
BCa upper	1.14 to 1.16

R output: `lm(formula = Debt2007 ~ Debt2006, data = debt)`

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.11113    3.60162   0.309     0.76
Debt2006     1.11594    0.01808  61.727   <2e-16

Residual standard error: 11.09 on 22 degrees of freedom
Multiple R-squared: 0.9943,     Adjusted R-squared: 0.994
F-statistic:  3810 on 1 and 22 DF,  p-value: < 2.2e-16
```

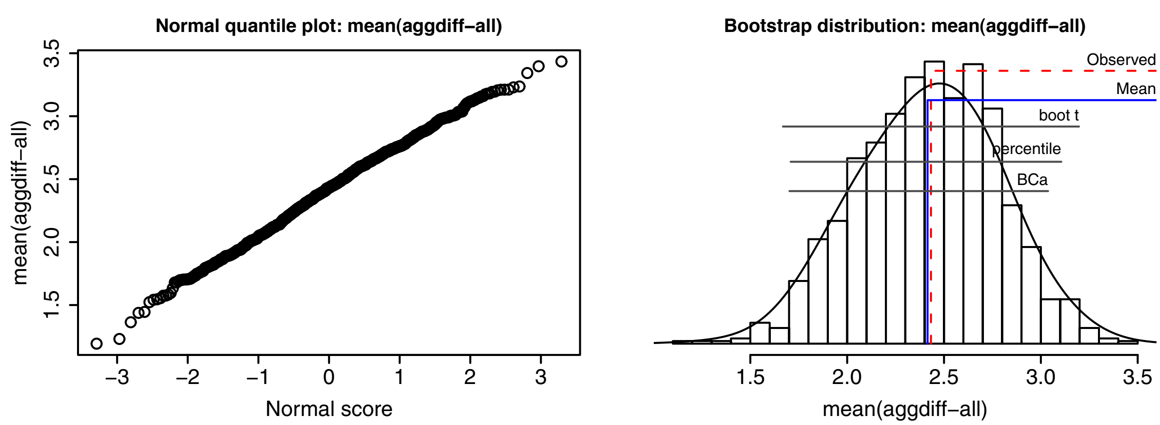


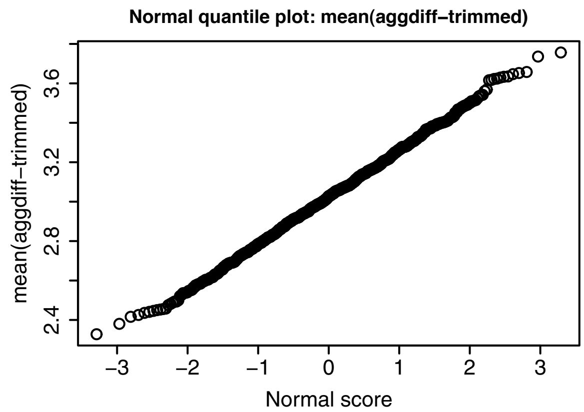


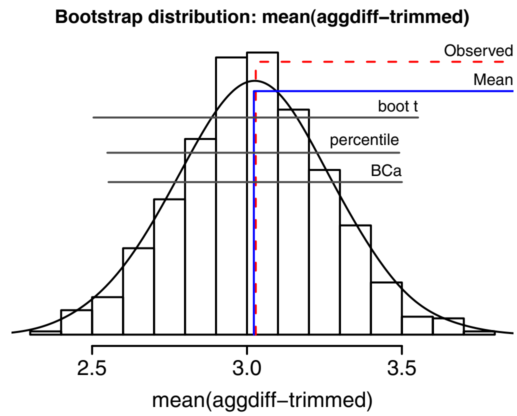
16.50. (a) The bootstrap distribution typically has slight left skewness for the full data set, and is close to Normal with the outliers removed. The bias is close to zero in each case, but SE_{boot} is substantially larger for the full data set. **(b)** Ranges for the BCa intervals are given on the right; the interval is higher and much narrower with the outliers excluded.

Typical ranges	
Bias (all points)	−0.04 to 0.04
Bias (trimmed)	−0.02 to 0.03
SE_{boot} (all points)	0.33 to 0.39
SE_{boot} (trimmed)	0.22 to 0.26
BCa lower (all)	1.42 to 1.76
BCa upper (all)	2.99 to 3.19
BCa lower (trimmed)	2.47 to 2.64
BCa upper (trimmed)	3.44 to 3.63

Note: *The outliers have the same effect on the standard t interval: With all points, the interval is 1.6240 to 3.2414, and with the outliers removed, the interval is 2.4679 to 3.5888.*







16.51. No, because we believe that one population has a smaller spread, but in order to pool the data, the permutation test requires that both populations be the same when H_0 is true.

16.52. The standard error is about $\sqrt{(0.04)(0.96)/250} \doteq 0.0124$. We should not feel comfortable declaring this to be significant at the 5% level, because 0.04 is less than one SE below 0.05.

16.53. **(a)** The observed difference in means is $\frac{57+53}{2} - 19 + 37 + 41 + 424 = 20.25$.
(b) Student results will vary, but should be one of the 15 (equally likely) possible values:

$$-20.25,\ -17.25,\ -16.5,\ -8.25,\ -5.25,\ -3.75,\ -3,\ 0,\ 5.25,\ 8.25,\ 9,\ 11.25,\ 12,\ 20.25$$

(c) The histogram shape will vary considerably. **(d)** Out of 20 resamples, the number which yield a difference of 20.25 or more has a binomial distribution with $n = 20$ and $p = 1/15$, so most students should get between 0 and 4, for a P-value between 0 and 0.2. **(e)** As was noted in part (b), only one resample gives a difference of means greater than or equal to the observed value, so the exact P-value is $1/15 \doteq 0.0667$.

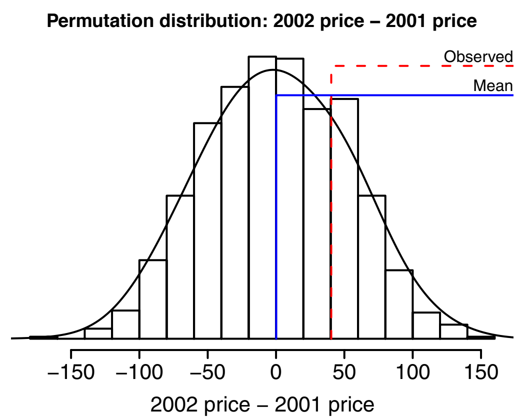
Note: *To determine the 15 possible values, note that the six numbers sum to 249. If the first two numbers add up to T, then the other four will add up to $249 - T$, and the difference in means will be $\frac{1}{2}T - \frac{1}{4}(249 - T) = \frac{3}{4}T - 62.25$. The values of T range from $19 + 37 = 56$ to $57 + 53 = 110$.*

Typical ranges	
Bias	-5 to 5
SE_{boot}	45 to 55
BCa lower	-51 to -27
BCa upper	137 to 201

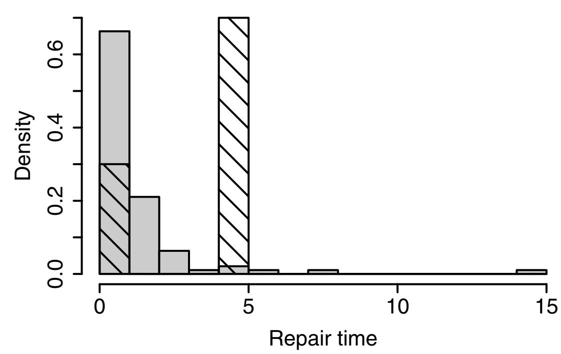
16.54. **(a)** We test H_0: $\mu_1 = \mu_2$ versus H_a: $\mu_1 \neq \mu_2$, where μ_1 is the mean selling price for all Seattle real estate transactions in 2001, and μ_2 is the mean selling price for Seattle real estate transactions in 2002. **(b)** With $\bar{x}_1 = 288.94$ and $\bar{x}_2 = 329.34$, we find $t \doteq 0.81$, df $\doteq 71.9$, and $P \doteq 0.4223$. (If we assume equal standard deviations, df $= 98$ and $P \doteq 0.4216$.) **(c)** With 1000 resamples, the two-sided P-value will typically be between 0.34 and 0.50. These are reasonably consistent with the P-value from part (b), leading to the same conclusion: There is little evidence that μ_1 and μ_2 differ. **(d)** Ranges for the BCa interval are on the right. These intervals include 0 and suggests that the two means are not significantly different at the 0.05 level, which is consistent with the conclusions in parts (b) and (c).

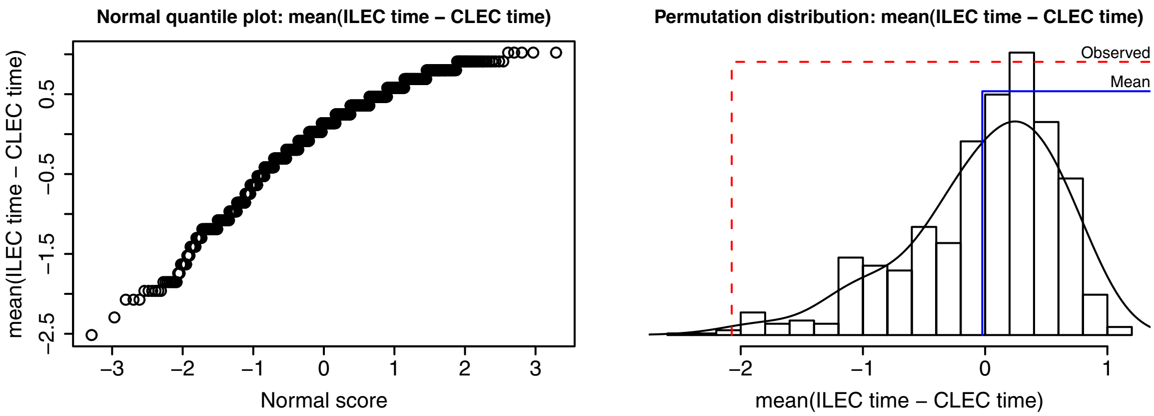
Note: *For part (c), note that if the true one-sided P-value is 0.21, then nearly all estimated one-sided P-values will be in the range* $0.21 \pm 3\sqrt{(0.21)(0.79)/1000} \doteq 0.21 \pm 0.04$*, so most estimated two-sided P-values will be between 0.34 and 0.50. That's a wide range; for more accuracy, use more resamples.*

For part (d), the lower and upper endpoints of the BCa intervals can vary quite a bit. When computing those intervals, R warns "Some BCa intervals may be unstable."



16.55. **(a)** The ILEC distribution (gray bars) is clearly skewed to the right, while the CLEC distribution is skewed to the left (although with only 10 observations, that might not mean anything). **(b)** In keeping with the discussion in Example 16.13, we use a one-sided alternative. For a test of H_0: $\mu_{\text{ILEC}} = \mu_{\text{CLEC}}$ versus H_a: $\mu_{\text{ILEC}} < \mu_{\text{CLEC}}$, we find $t \doteq -3.25$, df $\doteq 10.71$, and $P \doteq 0.004$. **(c)** Based on 1000 resamples (each of size 1000), the P-value typically ranges from 0.001 to 0.010. The permutation test does not require Normal distributions, and gives more accurate answers in the case of skewness. A plot of the permutation distribution shows there is substantial skewness. **(d)** The difference is significant at the 5% level (and usually at the 1% level).





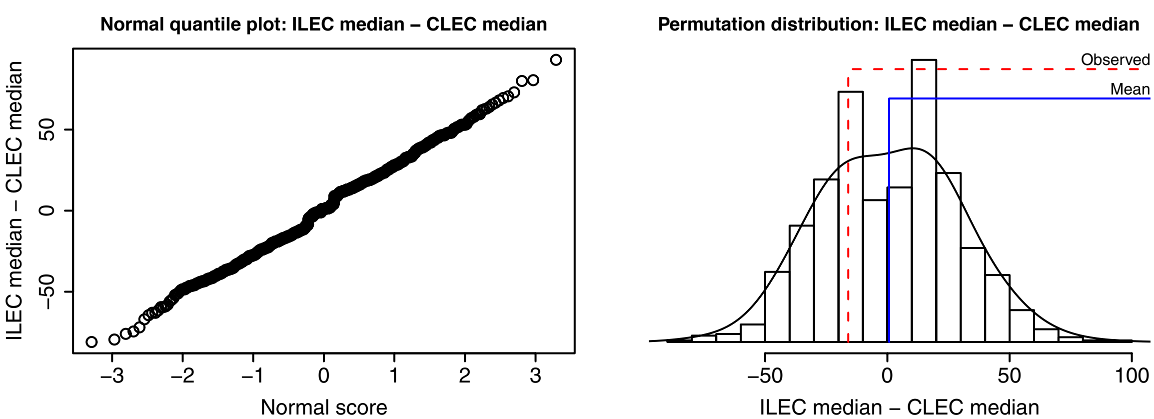
16.56. The standard deviations are approximately $\sqrt{P(1-P)/n}$, giving the results on the right.

Study	n	P	Standard deviation
DRP	1000	0.015	0.00384
Verizon	500,000	0.0183	0.00019

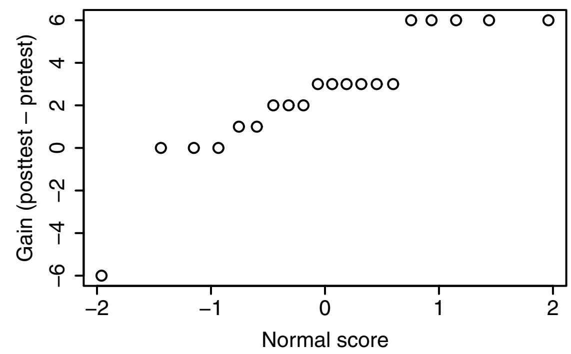
16.57. (a) The two populations should be the same shape, but skewed—or otherwise clearly non-Normal—so that the t test is not appropriate. **(b)** Either test is appropriate if the two populations are both Normal with the same standard deviation. **(c)** We can use a t test, but not a permutation test, if both populations are Normal with different standard deviations.

16.58. With resamples of size 1000, the P-value typically ranges from 0.53 to 0.68. If the price distributions in 2001 and 2002 were the same, then a difference in medians as large as we observed would occur more than half the time. We conclude that the difference is easily explained by chance, and is therefore not statistically significant.

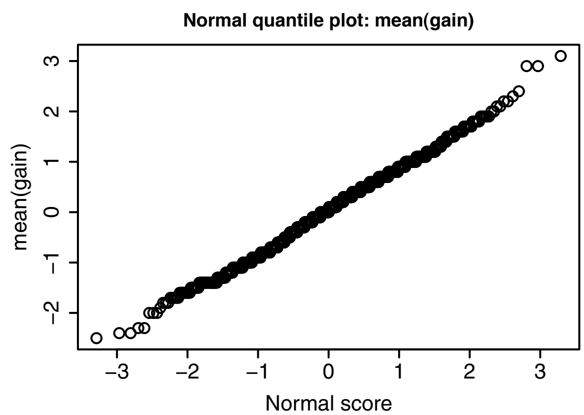
Note: *The permutation distribution (below) appears to be bimodal, but that does not affect our conclusion.*

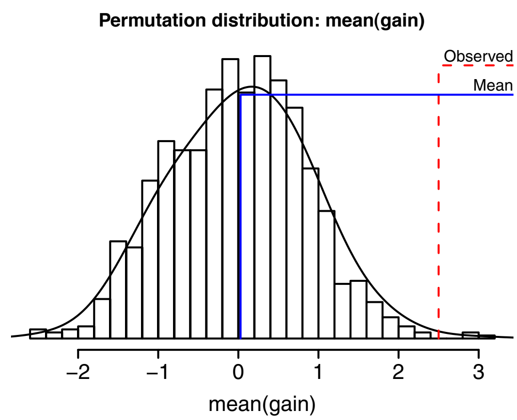


16.59. (a) We test H_0: $\mu = 0$ versus H_a: $\mu > 0$, where μ is the population mean difference before and after the summer language institute. We find $t \doteq 3.86$, df $= 19$, and $P \doteq 0.0005$. **(b)** The quantile plot (right) looks odd because we have a small sample, and all differences are integers. **(c)** The P-value is almost always less than 0.002. Both tests lead to the same conclusion: The difference is statistically significant.

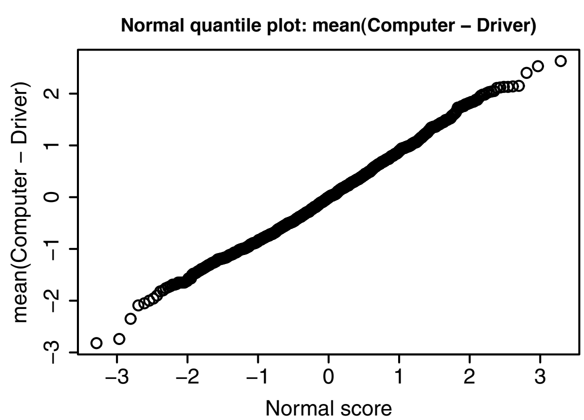


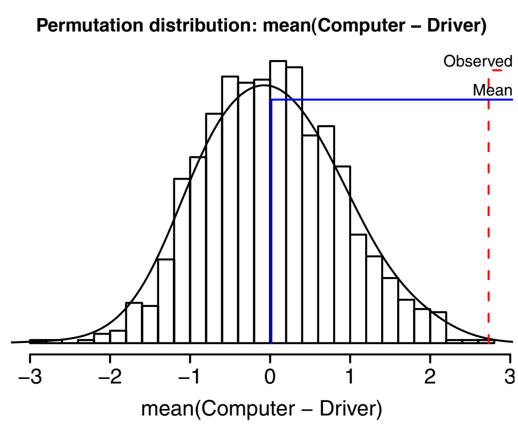
Note: *The text states that "the histogram [of the permutation distribution] looks a bit odd." In fact, different software produces different default histograms, some of which look fine. (This statement was made about the default histogram produced by S-PLUS.) To avoid potential confusion, check what your software does, and (if necessary) tell students to ignore that part of the question.*



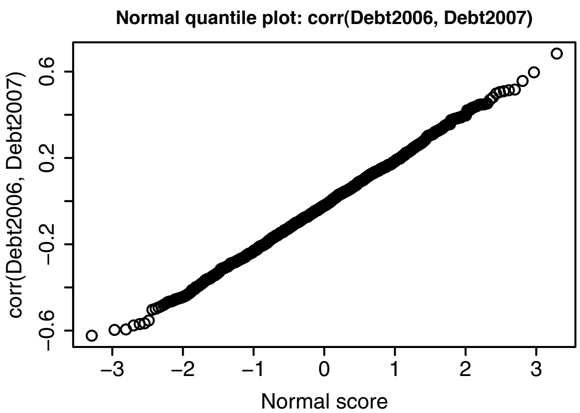


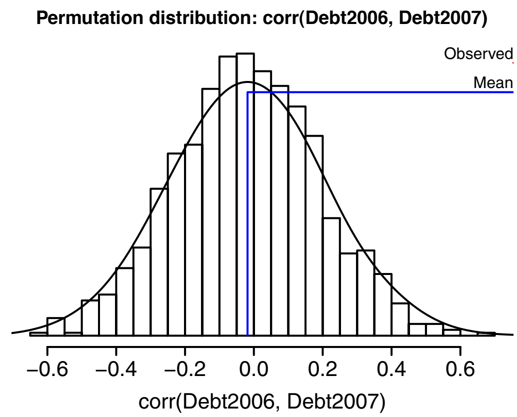
16.60. (a) We test H_0: $\mu_D = \mu_C$ versus H_a: $\mu_D < \mu_C$, where μ_D is the mean driver-calculated mpg, and μ_C is the mean computer mpg. We find $t \doteq 4.36$, df $= 19$, and $P \doteq 0.0002$. This is very strong evidence against the null hypothesis. **(b)** The permutation test P-value is almost always 0.001 or less. This is reasonably close to the value from the t test.



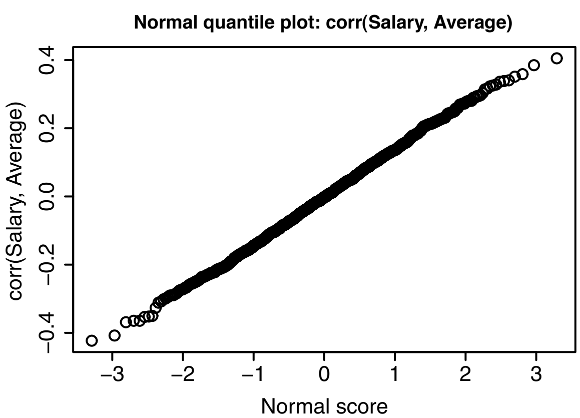


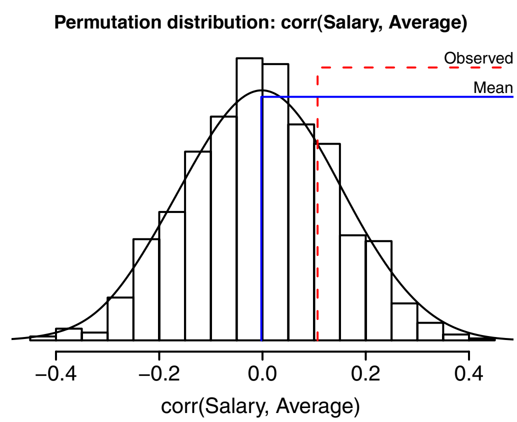
16.61. We test H_0: $\rho = 0$ versus H_a: $\rho \neq 0$, where ρ is the population correlation. (One could also justify the one-sided alternative $\rho > 0$ in this case; the ultimate conclusion is the same for either alternative.) The permutation distribution (found by permuting the debts from one year, then computing the correlation) is roughly Normal. In the histogram of the permutation distribution below, the observed correlation ($r \doteq 0.997$) is not marked because it lies far out on the high tail, nearly five standard deviations above the mean (0). Consequently, the reported P-value is nearly always 0, confirming the very strong evidence found in the solution to Exercise 16.45.



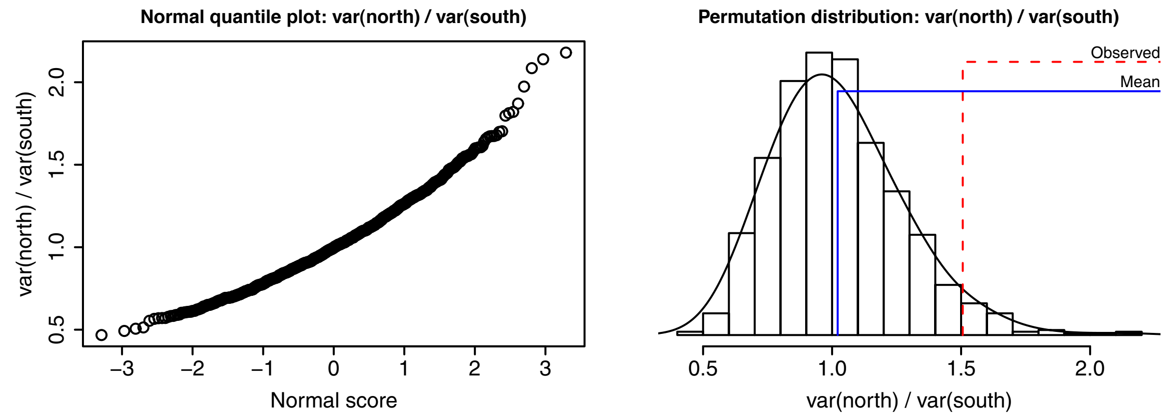


16.62. **(a)** We test H_0: $\rho = 0$ versus H_a: $\rho > 0$, where ρ is the population correlation. **(b)** With 10,000 resamples, the P-value from the permutation test will almost always be between 0.21 and 0.25. This does not provide significant evidence against the null hypothesis. (With fewer resamples, the P-values have a wider range, but any reasonable resample size will lead to the same conclusion.)



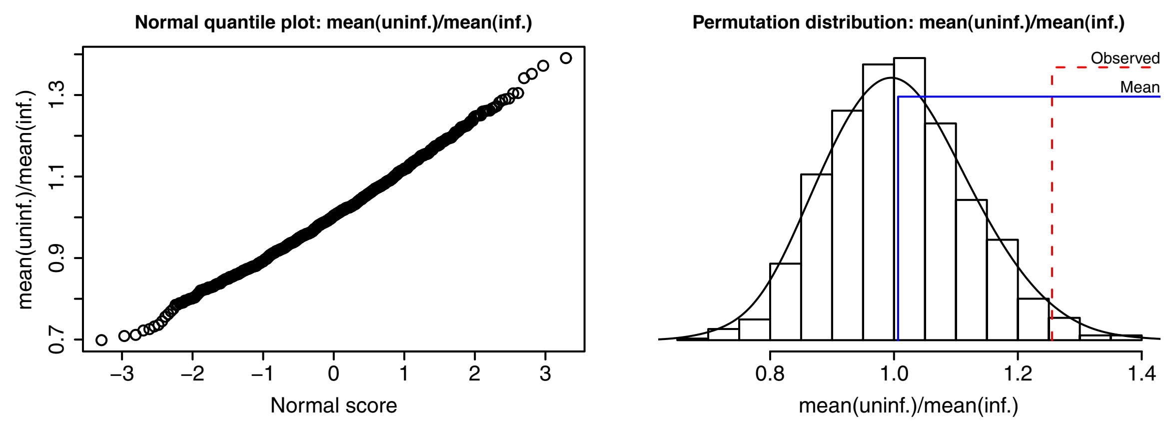


16.63. For testing H_0: $\sigma_1 = \sigma_2$ versus H_a: $\sigma_1 \neq \sigma_2$, the permutation test P-value will almost always be between 0.065 and 0.095. In the solution to Exercise 7.105, we found $F \doteq 1.50$ with df $= 29$ and 29, for which $P \doteq 0.2757$—three or four times as large. In this case, the permutation test P-value is smaller, which is typical of short-tailed distributions.



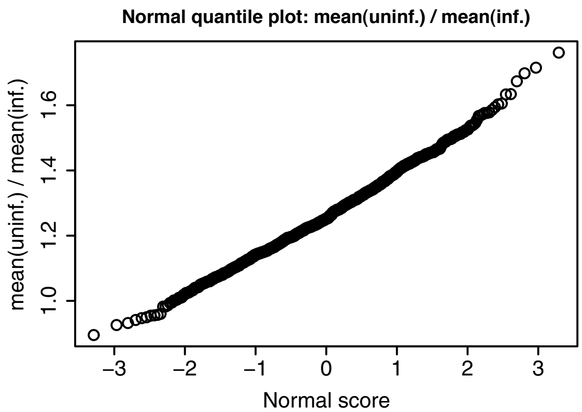
16.64. (a) The group means (in μmol/L) are $\bar{x}_1 \doteq 0.778$ (group 1, uninfected) and $\bar{x}_2 \doteq 0.620$ (group 2, infected). **(b)** If $\mathcal{R} = \mu_1/\mu_2$ is the ratio of the population means, we test H_0: $\mathcal{R} = 1$ versus H_a: $\mathcal{R} > 1$. The one-sided P-value is typically between 0.014 and 0.21. The permutation distribution (below) is centered near 1 with standard deviation approximately 0.11; it is roughly Normal with some right skewness. We expect the permutation distribution to be centered at about 1, because that is the null hypothesis value for the ratio.

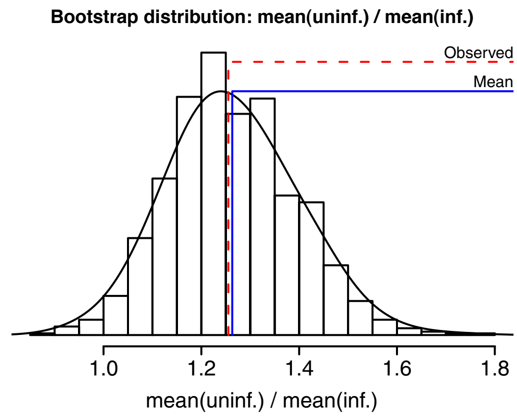
Note: *Our permutation resampling will, on the average, produce $\bar{x}_1 = \bar{x}_2$, so it seems "obvious" that the ratio $\bar{x}_1/\bar{x}_2$ should equal 1 on the average. Of course, one should beware of accepting the "obvious"; in general, the expected value of a ratio is* not *equal to the ratio of the expected values, although it will often (as in this case) be close.*



16.65. For the permutation test, we must resample in a way that is consistent with the null hypothesis. Hence we pool the data—assuming that the two populations are the same—and draw samples (without replacement) for each group from the pooled data. For the bootstrap, we do not assume that the two populations are the same, so we sample (with replacement) from each of the two datasets separately, rather than pooling the data first.

Note: *Shown below is the bootstrap distribution for the ratio of means; comparing this with the permutation distribution from the previous solution illustrates the effect of the resampling method.*



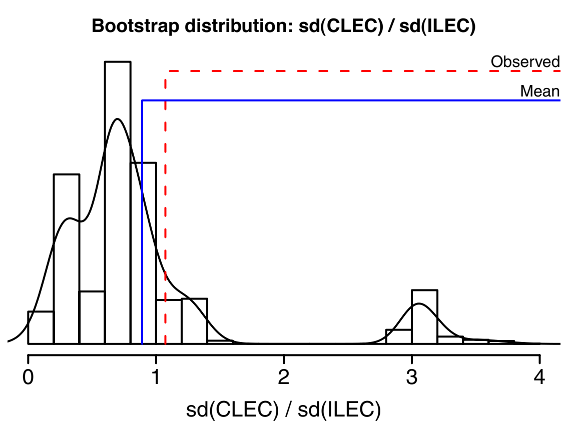


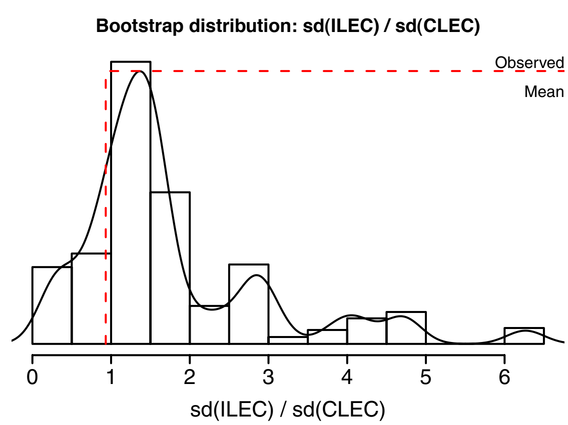
16.66. **(a)** Ranges for the BCa interval (based on 1000 resamples) are given on the right. Note that some of these intervals include 1, suggesting that for some resamples, we could not reject H_0: $\mathfrak{R} = 1$. **(b)** The bootstrap distribution (shown with the solution to the previous problem) is right skewed, with relatively little bias. Typically, the percentile interval is shifted slightly to the right of the BCa interval, which means that it suggests slightly stronger evidence against H_0: $\mathfrak{R} = 1$.

Typical ranges	
Bias	−0.01 to 0.02
SE_{boot}	0.12 to 0.15
BCa lower	0.97 to 1.06
BCa upper	1.47 to 1.58
Percentile lower	0.99 to 1.05
Percentile upper	1.49 to 1.58

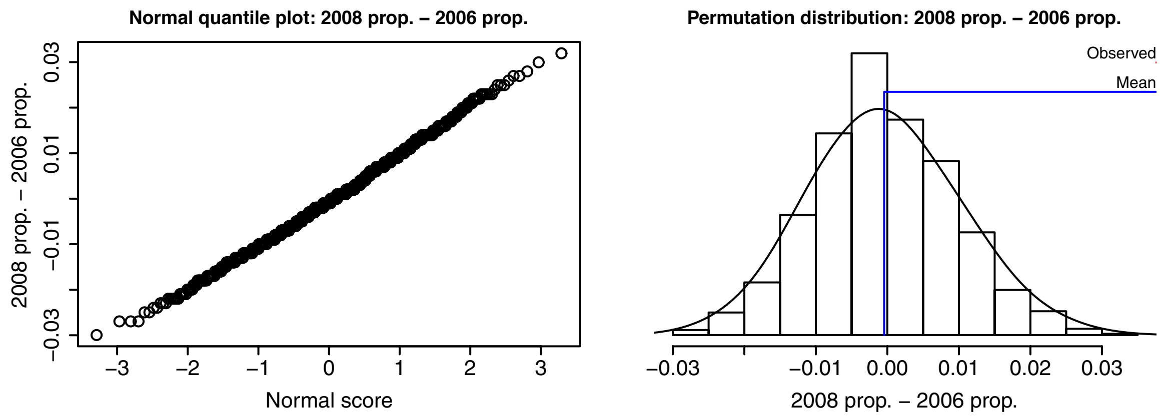
Note: *If students use a larger resample size, the intervals show less variability. with 10,000 resamples, both the BCa and percentile intervals will almost certainly (barely) not include 1; in particular, the BCa lower bound is typically between 1.003 and 1.023.*

16.67. **(a)** The resampled CLEC standard deviation is sometimes 0, so for best results (that is, to avoid infinite ratios), put that standard deviation in the numerator. Both bootstrap distributions are shown below. (We do not need quantile plots to confirm that these distributions are non-Normal.) Regardless of which ratio we use, the resulting P-value is close to 0.37. **(b)** The difference in the P-values is evidence of the inaccuracy of the F test; these distributions clearly do not satisfy the Normality assumption.



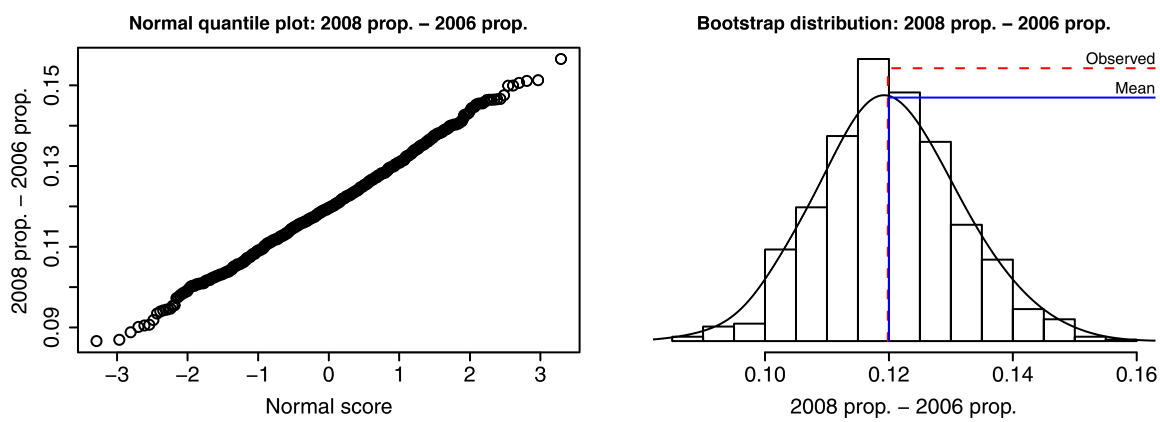


16.68. To test H_0: $p_1 = p_2$ versus H_a: $p_1 \neq p_2$, we resample (without replacement) from the pooled data: $2822 + 1553 = 4375$ subjects, of which $198 + 295 = 493$ have downloaded a podcast. The exact approach depends on the software used; one way is to code each of the 4375 responses as 0 or 1, and do a permutation test for a difference of means. Regardless of the approach, the result should be the same: The observed difference ($\hat{p}_2 - \hat{p}_1 \doteq 0.1198$) is about 12 standard deviations above the mean, and the reported P-value is nearly always 0.



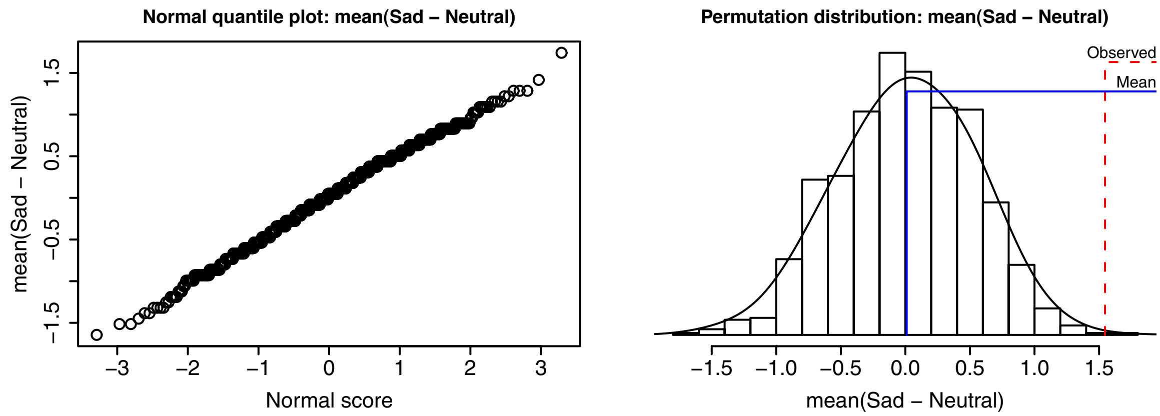
16.69. The bootstrap distribution looks quite Normal, and (as a consequence) all of the bootstrap confidence intervals are similar to each other, and also are similar to the standard (large-sample) confidence interval: 0.0981 to 0.1415.

Note: *At the time these solutions were written, R's bootstrapping package would fail if asked to find the BCa confidence interval for this exercise.*

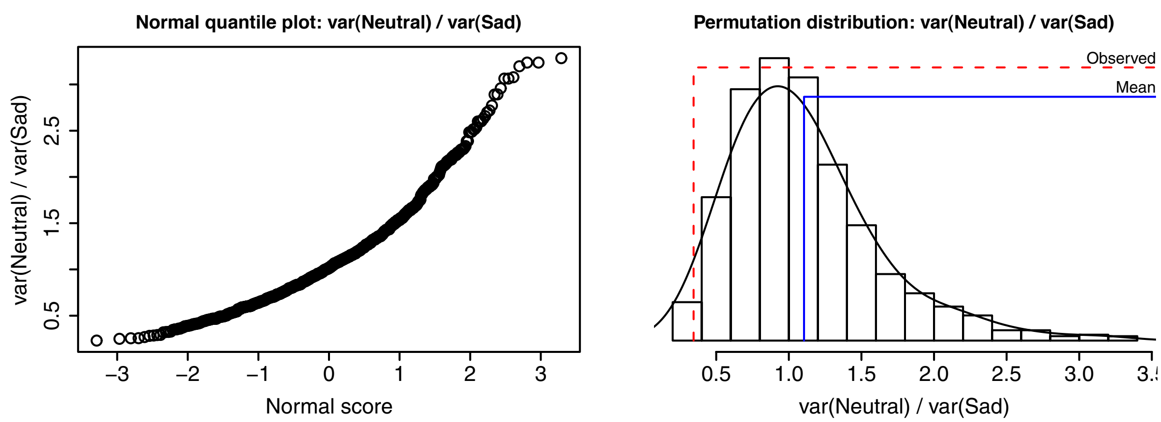


16.70. See the solution to Exercise 7.65 for stemplots and summary statistics, and the unpooled test; the solution to Exercise 7.86 gives the details for the pooled test. In Exercise 7.65, the text suggested that there was a prior suspicion that that sad group would be willing to pay more, so we used a one-sided alternative. Here we use a two-sided alternative. **(a)** The unpooled t statistic is $t \doteq -4.30$ with df $\doteq 26.48$, for which the two-sided P-value is $P \doteq 0.0002$. **(b)** The pooled test gives $t \doteq -4.10$, df $= 2$, and $P \doteq 0.0003$. **(c)** Apart from some granularity (visible in the quantile plot), the permutation distribution is reasonably Normal. The observed difference is about three standard deviations above the mean, and $P < 0.002$ (nearly always). **(d)** Student preferences will vary. Note that the permutation test is safest because it makes the fewest assumptions about the populations, while the pooled

t test makes the most assumptions. Given the identical conclusions and the similarity of strength of the evidence, there is little reason to have a strong preference here.



16.71. **(a)** The standard test of H_0: $\sigma_1 = \sigma_2$ versus H_a: $\sigma_1 \neq \sigma_2$ leads to $F = 0.3443$ with df 13 and 16; $P \doteq 0.0587$. **(b)** The permutation P-value is typically between 0.02 and 0.03. **(c)** The P-values are similar, even though technically, the permutation test is significant at the 5% level, while the standard test is (barely) not. Because the samples are too small to assess Normality, the permutation test is safer. (In fact, the population distributions are discrete, so they cannot follow Normal distributions.)

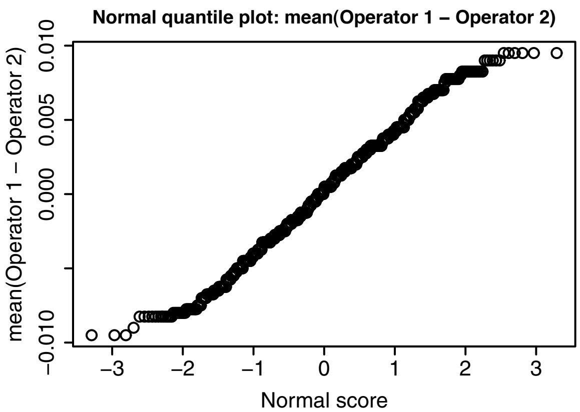


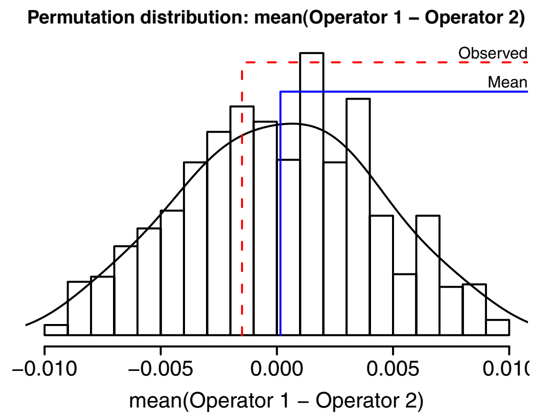
16.72. To compare the means, we need a matched-pairs permutation test, which gives a P-value near 0.78—no reason to suspect a systematic difference in the operators' measurements.

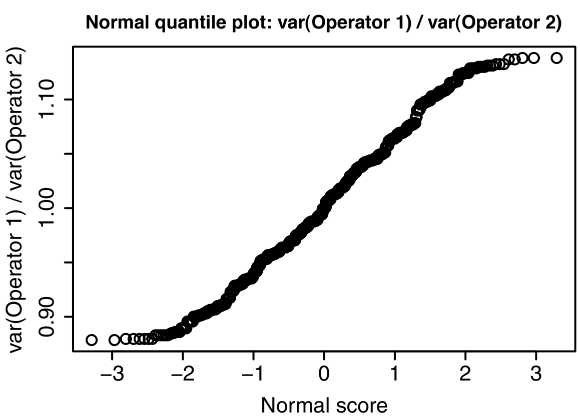
There is not really a legitimate way to compare the spreads with the data we have. Most of the variation in each operator's measurements can be attributed to variation in the subjects being measured, rather than variation due to the operator's abilities. Nonetheless, we can do this comparison by randomly swapping (or not) observations within each matched pair, and then examining the ratio of variances (or equivalently, the standard deviations). The permutation distribution of the variance ratio is given below; it has $P \doteq 0.66$. In both cases there is not statistically significant evidence of a difference between the operators. The differences could easily arise by chance; even larger differences would occur by chance more than half the time.

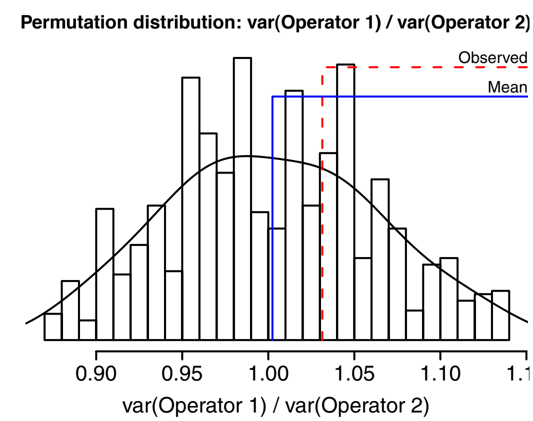
Note: *For a legitimate comparison of the spreads for the two operators, we would want* *to have multiple measurements on each subject.*

Although the ratio of variances is the most common comparison, we could also compute the difference *between the variances (or standard deviations). These approaches yield P-values similar to those found for the ratio.*



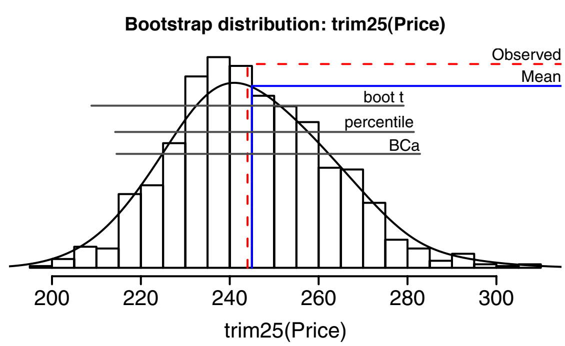






16.73. See the solution to Exercise 16.18 for another view of the bootstrap distribution. Ranges for the bootstrap t, percentile, and BCa intervals are given on the right. All have similar upper endpoints, but the lower endpoint for the bootstrap t is typically less than the others (because it ignores the skewness in the bootstrap distribution). It appears that any of the other intervals—including the percentile interval—would be more reliable.

Typical ranges	
t lower	205.5 to 211.5
t upper	276.5 to 282.5
Percentile lower	208.4 to 216.5
Percentile upper	276.2 to 288.8
BCa lower	207.9 to 217.4
BCa upper	276.0 to 293.6



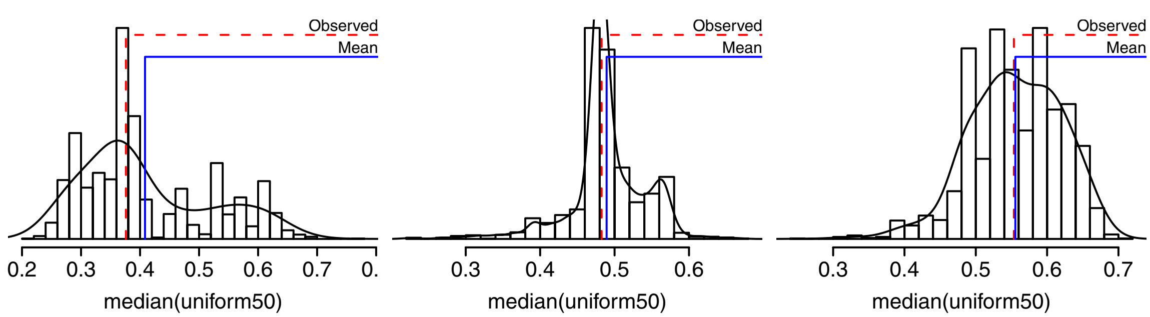
16.74. See the solution to Exercise 16.5 for the bootstrap distribution. **(a)** The bootstrap distribution is approximately Normal, with a very slight right skew. A t interval should be acceptable unless high accuracy is needed. **(b)** Ranges for the t and BCa intervals are given on the right. **(c)** The BCa interval is typically shifted to the right of the t interval, as we might expect because of the slight right skew.

Typical	ranges
t lower	3.1 to 3.4
t upper	5.8 to 6.1
BCa lower	3.2 to 3.6
BCa upper	5.9 to 6.5

16.75. All answers (including the shape of the bootstrap distribution) will depend strongly on the initial sample of uniform random numbers. The median M of these initial samples will be between about 0.36 and 0.64 about 95% of the time; this is the center of the bootstrap t confidence interval. **(a)** For a uniform distribution on 0 to 1, the population median is 0.5. Most of the time, the bootstrap distribution is quite non-Normal; three examples are shown below. **(b)** SE_{boot} typically ranges from about 0.04 to 0.12 (but may vary more than that, depending on the original sample). The bootstrap t interval is therefore roughly $M \pm 2SE_{boot}$. **(c)** The more sophisticated BCa and tilting intervals may or may not be similar to the bootstrap t interval. The t interval is not appropriate because of the non-Normal shape of the bootstrap distribution, and because SE_{boot} is unreliable for the sample median (it depends strongly on the sizes of the gaps between the observations near the middle).

Note: *Based on 5000 simulations of this exercise, the bootstrap t interval $M \pm 2SE_{boot}$ will capture the true median (0.5) only about 94% of the time (so it slightly underperforms its intended 95% confidence level). In the same test, both the percentile and BCa intervals included 0.5 over 95% of the time, while at the same time being narrower than the bootstrap t interval nearly two-thirds of the time. These two measures (achieved confidence level, and width of confidence interval) both confirm the superiority of the other intervals.*

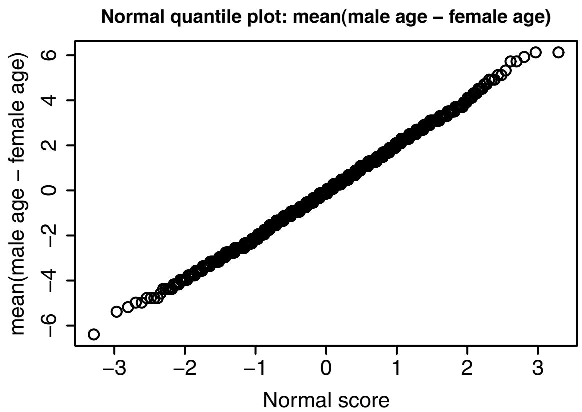
The bootstrap percentile, BCa, and tilting intervals do fairly well despite the high variability in the shape of the bootstrap distribution. They give answers similar to the exact rank-based confidence intervals obtained by inverting hypothesis tests. One variation of tilting intervals matches the exact intervals.

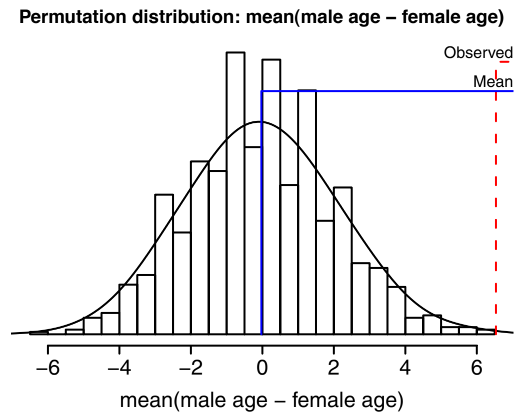


16.76. **(a)** The difference appears to be quite large; it should be significant. **(b)** The two-sided permutation P-value is about 0.000655. With 1000 resamples, students will typically get a P-value of no more than 0.004. **(c)** We conclude that there is significant evidence that the mean ages differ. The t test P-value is similar to the (true) permutation P-value, although student estimates of the latter might be too high.

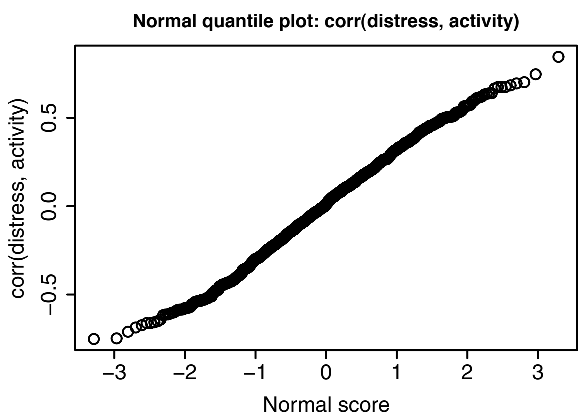
Male		Female
	1	9
	2	01
3	2	2233
5	2	4
6	2	
899	2	9
01	3	
22	3	
5	3	

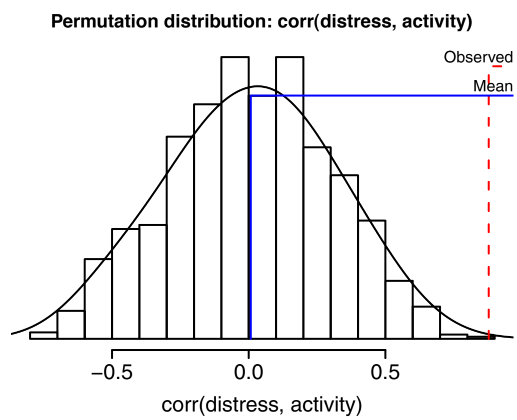
Note: *Some software will compute the exact permutation test P-value; it is* $\frac{110}{167{,}960} \doteq 0.000655$*. This is about three times larger than the standard t test result:* $P \doteq 0.000223$.





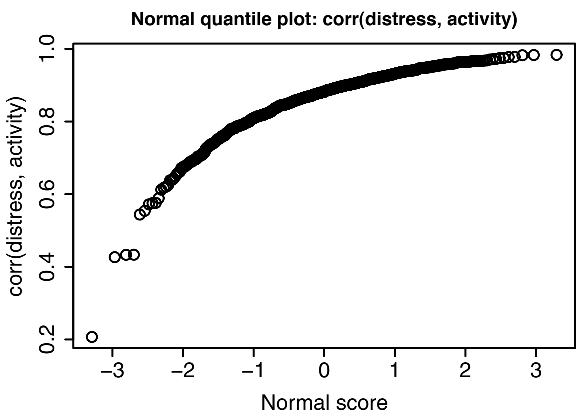
16.77. See the solution to Exercise 2.33 for a scatterplot. The permutation distribution (found by permuting one variable and computing the correlation) is roughly Normal, and the observed correlation ($r \doteq 0.878$) lies far out on the high tail, about three standard deviations above the mean (0). We conclude there is a significant positive relationship.

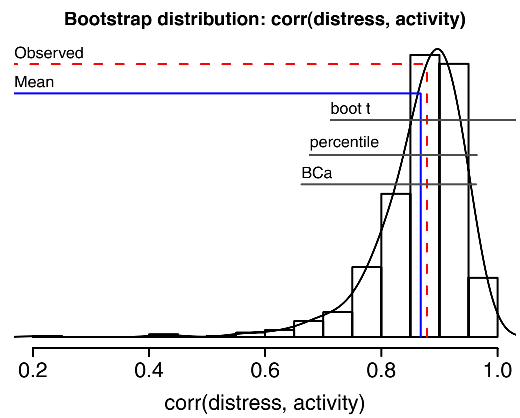




16.78. **(a)** The permutation distribution is centered near 0, because for a hypothesis test, we resample in a way that is consistent with H_0: $\rho = 0$. In contrast, the bootstrap is centered near the observed correlation of 0.878. The right tail is bounded above by 1, whereas the left tail can be much longer. **(b)** Ranges for the intervals are given on the right. (Because of the skew, the t interval is a poor choice for this situation, as evidenced by the upper limit exceeding 1.) None of the intervals are even close to including 0; we conclude that there is a significant positive relationship.

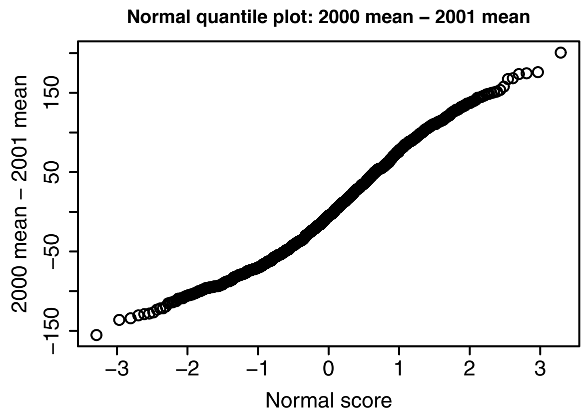
Typical ranges	
t lower	0.67 to 0.74
t upper	1.02 to 1.09
Percentile lower	0.61 to 0.71
Percentile upper	0.95 to 0.97
BCa lower	0.51 to 0.70
BCa upper	0.94 to 0.97

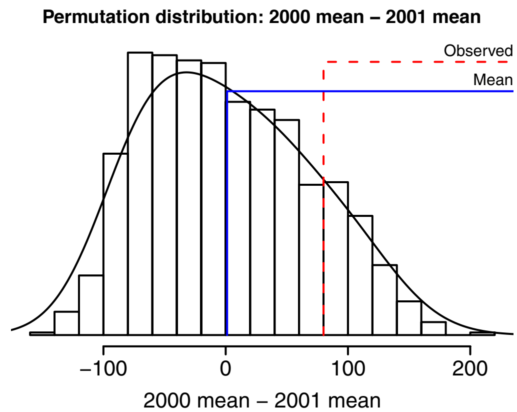




16.79. **(a)** The 2001 data is slightly skewed, but close to Normal given the sample size (50). The 2000 data is strongly right-skewed with two high outliers; a sample of size 20 is probably not enough to compensate. **(b)** The two-sided P-value for the permutation test is approximately 0.28. We conclude that there is not strong evidence that the mean selling prices are different for all Seattle real estate in 2000 and in 2001.

2000 prices		2001 prices	
1	3346899	0	5677
2	001488	1	0134445799
3	3669	2	0011123444677899
4	8	3	1123457
5		4	25556788
6		5	017
7		6	8
8		7	1
9			
10			
11	0		
12			
13			
14			
15			
16			
17			
18	4		



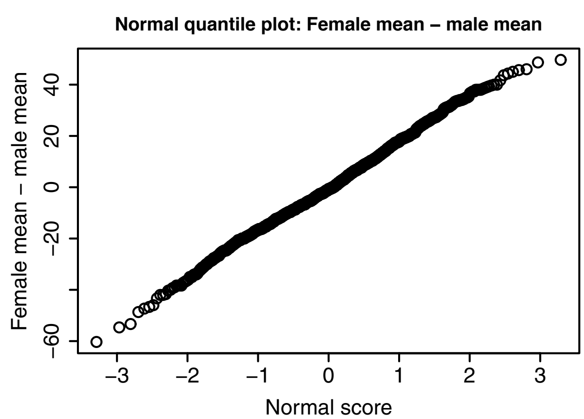


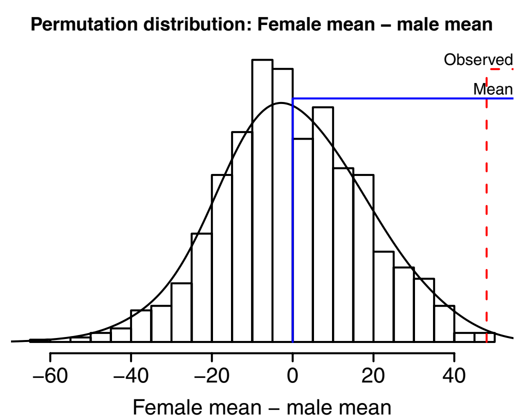
16.80. **(a)** See the solution to Exercise 1.41 for stemplots. Summary statistics (all in units of minutes):

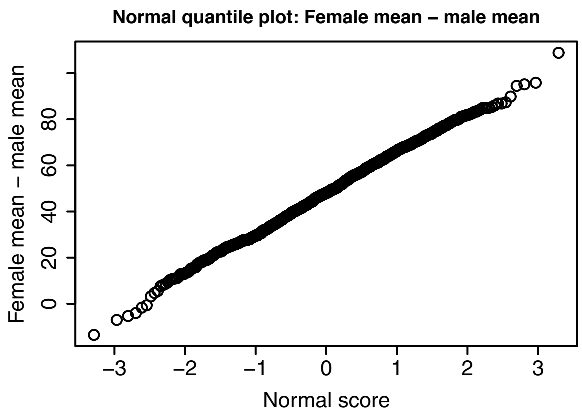
	$\bar{x}$	s	Min	Q_1	M	Q_3	Max
Men	117.2	74.24	0	60	120	150	300
Women	165.2	56.51	60	120	175	180	360

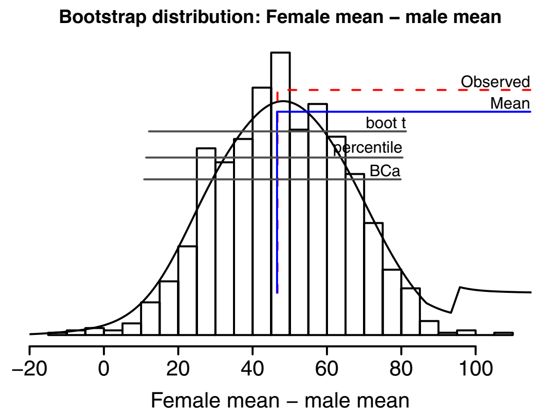
Typical ranges	
t lower	12.0 to 17.0
t upper	79.0 to 84.0
Percentile lower	8.8 to 18.9
Percentile upper	75.4 to 85.2
BCa lower	7.4 to 20.8
BCa upper	75.1 to 87.9

(b) The (unpooled) two-sample t test of H_0: $\mu_F = \mu_M$ versus H_a: $\mu_F \neq \mu_M$ gives $t \doteq 2.82$, df $\doteq 54.2$, and $P \doteq 0.0067$—a significant difference. A 95% confidence interval for the difference $\mu_F - \mu_M$ is 13.85 to 82.15 minutes. **(c)** A two-sided permutation test for the difference of means typically gives P no more than 0.02 (with 1000 resamples). The bootstrap distribution is slightly skewed; confidence intervals are similar to the standard t interval, although the percentile and BCa intervals are sometimes shifted to the left.



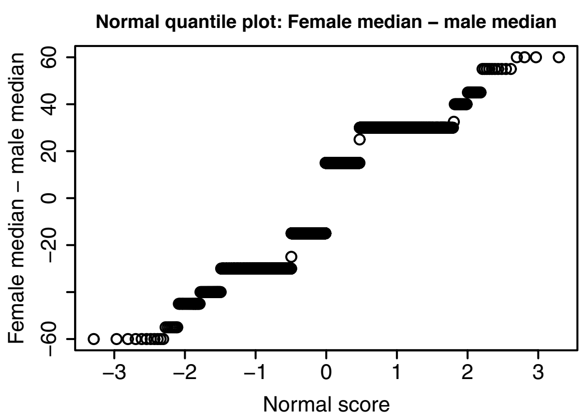


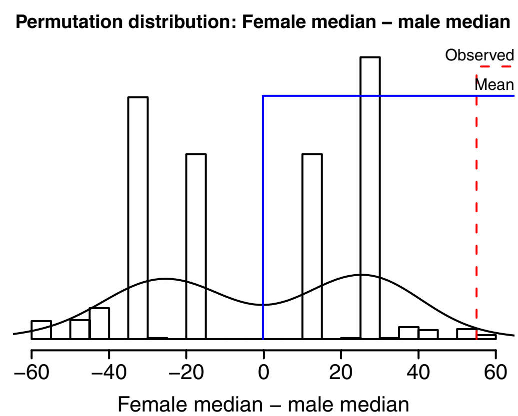


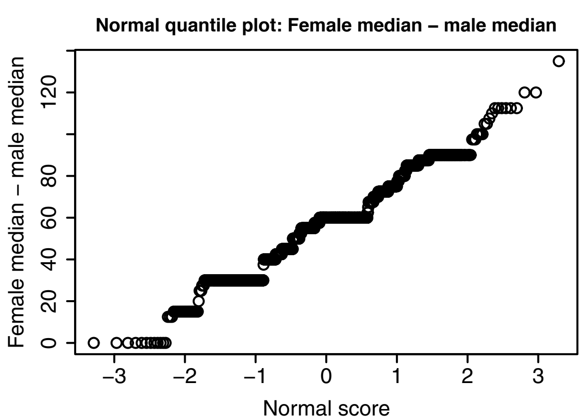


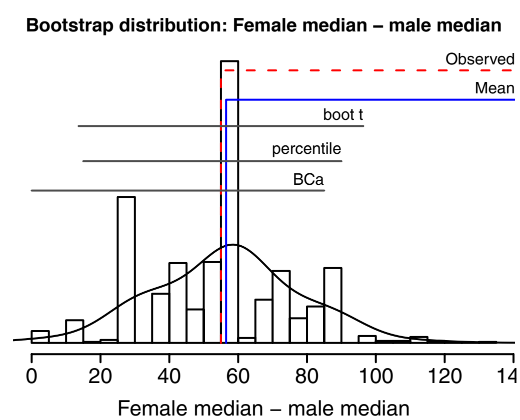
16.81. The permutation and bootstrap distributions for the difference in medians are extremely non-Normal, with many gaps and multiple peaks. In this situation, we have conflicting results: The permutation test gives fairly strong evidence of a difference (the two-sided P-value is roughly 0.032), but the BCa interval for the difference in medians nearly always includes 0.

Typical ranges	
t lower	9.6 to 16.5
t upper	93.5 to 100.4
Percentile lower	0 to 30
Percentile upper	90 to 100
BCa lower	−32.5 to 0
BCa upper	75 to 90



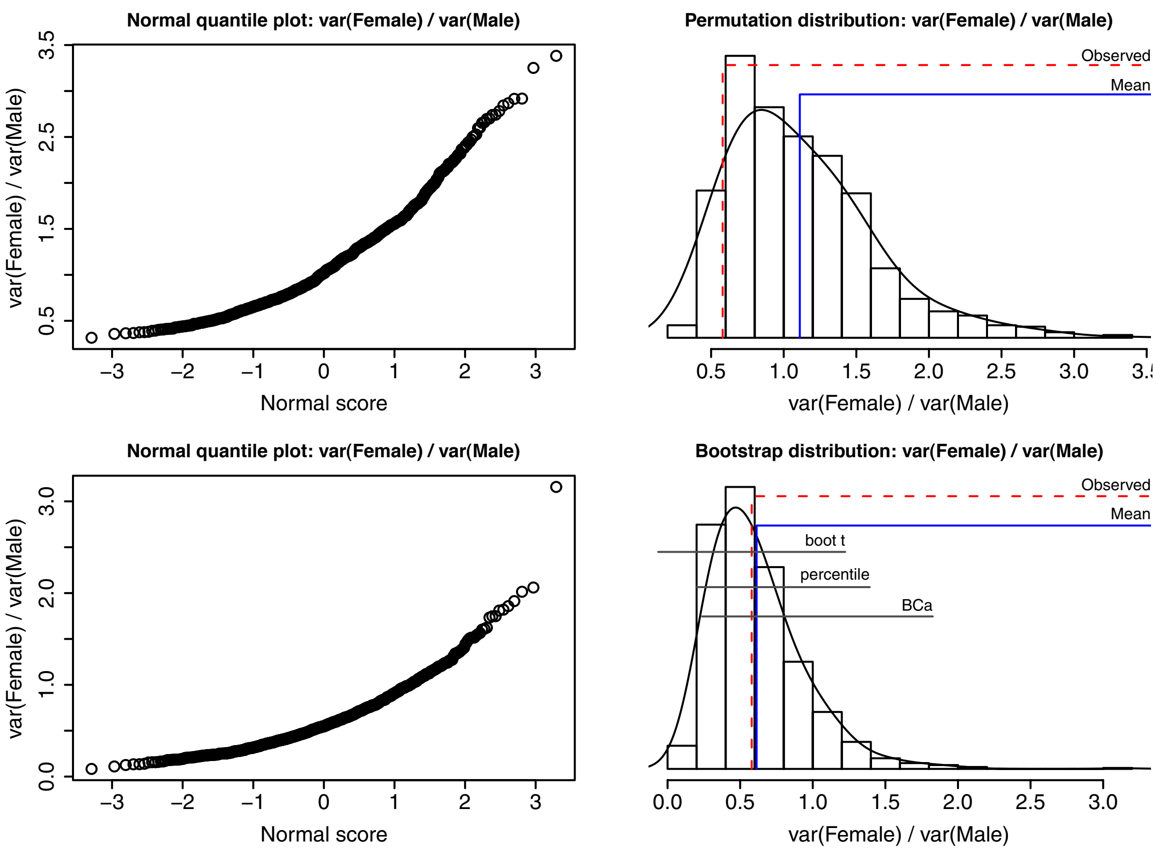






16.82. The standard test for equality of variances gives $F \doteq 0.58$ with df 29 and 29, for which $p = 0.1477$. **(a)** Using a permutation test, the two-sided P-value is about 0.226. Ranges for the bootstrap intervals are on the right; the bootstrap t is a bad choice for this sharply skewed distribution. **(b)** The variances are equal if and only if the standard deviations are equal. Any conclusions about the ratio of variances from the bootstrap and permutation distributions has an equivalent conclusion about the ratio of standard deviations. **(c)** We have strong evidence that the means of the two distributions are different, but cannot reject H_0: $\sigma_F = \sigma_M$. The evidence regarding the medians is mixed.

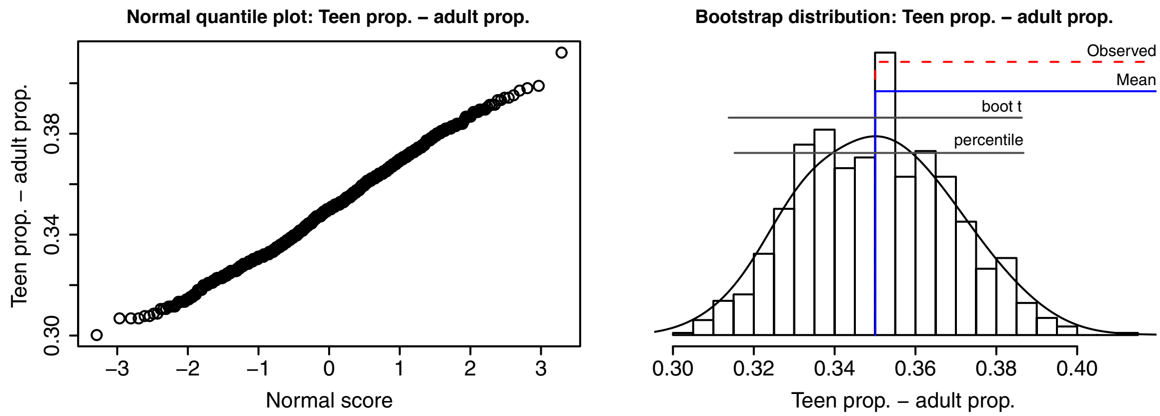
Typical ranges	
t lower	−0.22 to 0.02
t upper	1.14 to 1.38
Percentile lower	0.17 to 0.23
Percentile upper	1.24 to 1.62
BCa lower	0.21 to 0.27
BCa upper	1.42 to 2.42



16.83. See Exercise 8.55 for more details about this survey. The bootstrap distribution appears to be close to Normal; bootstrap intervals are similar to the large-sample interval (0.3146 to 0.3854).

Note: *At the time of this writing, R's bootstrapping package would not compute the BCa intervals for this exercise.*

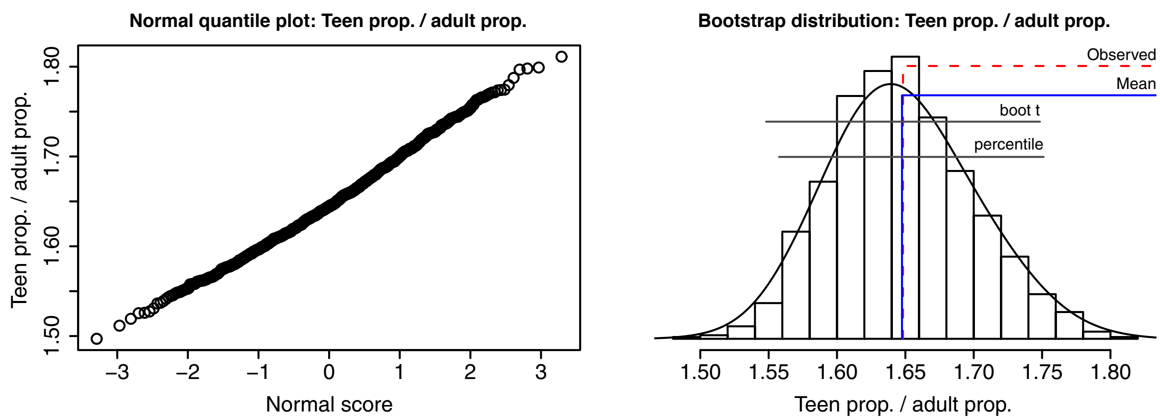
Typical ranges	
t lower	0.31 to 0.32
t upper	0.38 to 0.39
Percentile lower	0.30 to 0.32
Percentile upper	0.38 to 0.39



16.84. The bootstrap distribution is slightly skewed, but close enough to Normal that there is little difference among the interval methods.

Note: *At the time of this writing, R's bootstrapping package would not compute the BCa intervals for this exercise.*

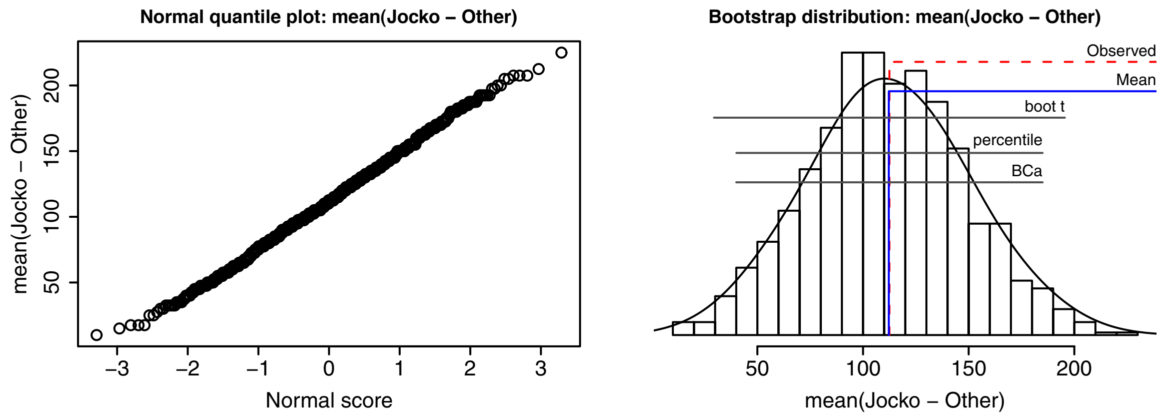
Typical ranges	
t lower	1.54 to 1.56
t upper	1.73 to 1.76
Percentile lower	1.54 to 1.57
Percentile upper	1.73 to 1.77



16.85. **(a)** This is the usual way of computing percent change: $89/54 - 1 = 0.65$. **(b)** Subtract 1 from the confidence interval found in Exercise 16.84; this typically gives an interval similar to 0.55 to 0.75. **(c)** Preferences will vary.

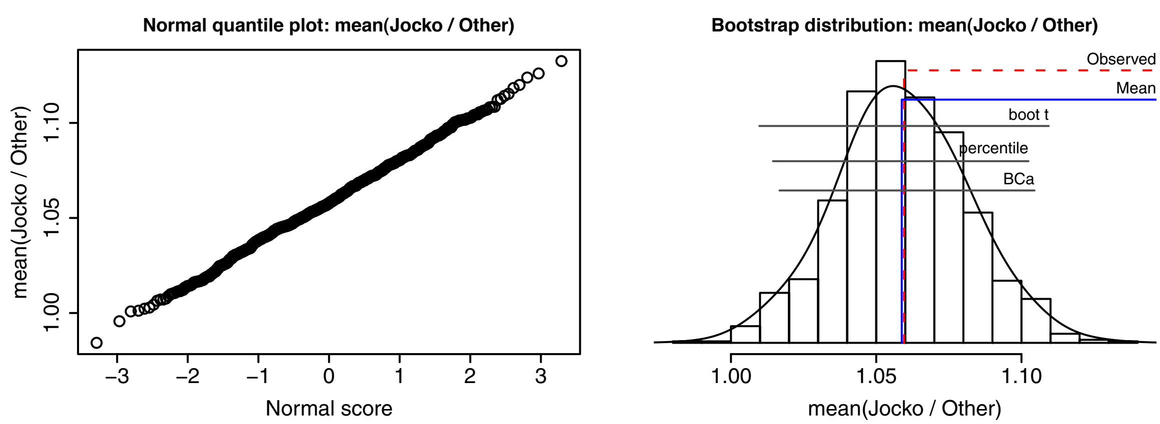
16.86. **(a)** Jocko's mean estimate is \$1827.5; the other garage's mean is \$1715. The matched-pairs t interval for the difference is $\$112.5 \pm \$88.52 = \$23.98$ to \$201.02. **(b)** Because these are matched pairs, we resample the differences. The distribution is reasonably close to Normal; ranges for the bootstrap intervals are on the right. **(c)** The bootstrap t interval is similar to the standard t interval; the other intervals are typically narrower.

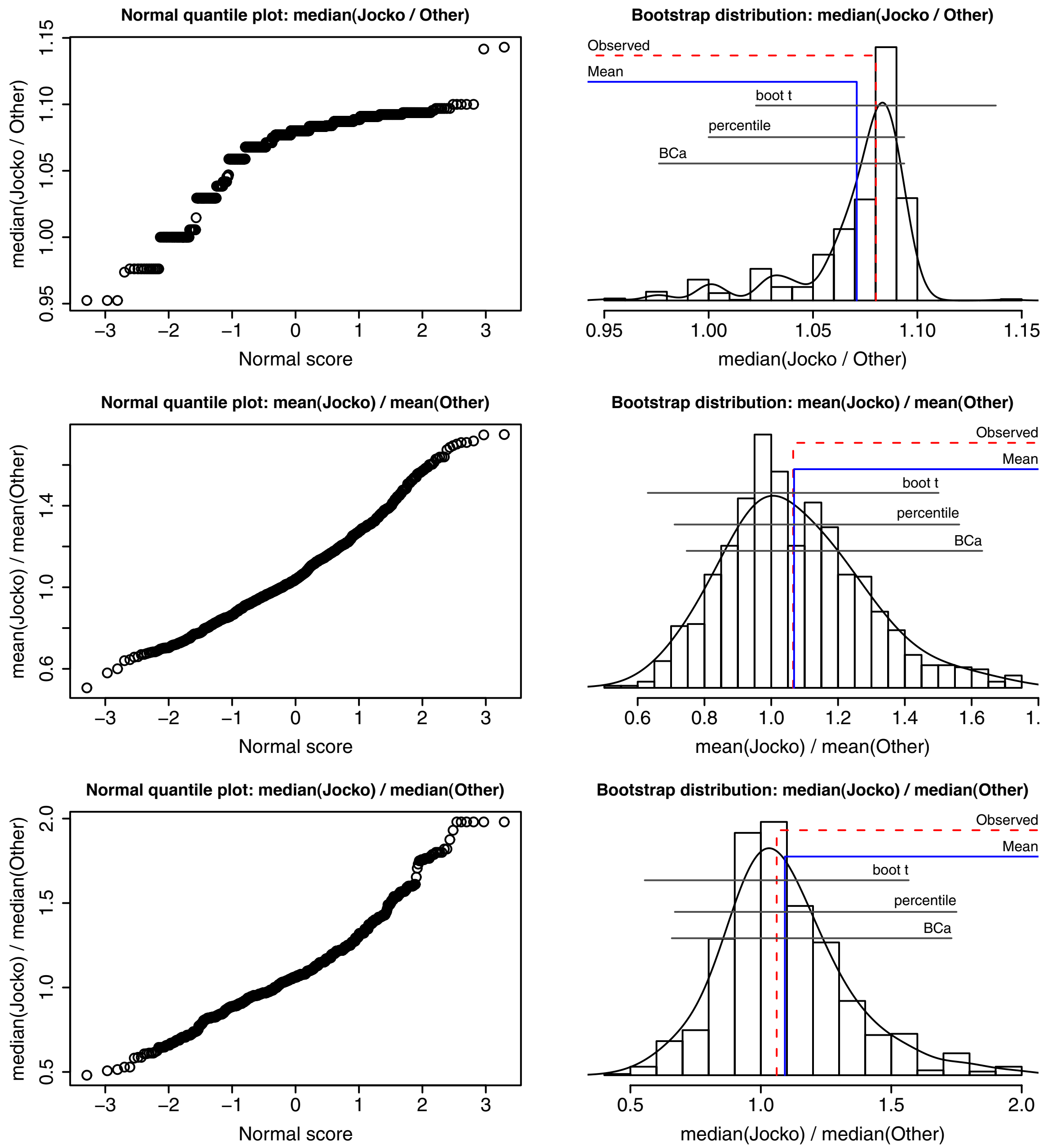
Typical ranges	
Bias	−3.7 to 3.7
SE_{boot}	34.7 to 39.9
t lower	22.3 to 34.0
t upper	191.0 to 202.7
Percentile lower	25.1 to 47.6
Percentile upper	175 to 195
BCa lower	20 to 45
BCa upper	170 to 195



16.87. **(a)** The mean ratio is 1.0596; the usual t interval is $1.0596 \pm (2.262)(0.02355) \doteq 1.0063$ to 1.1128. The bootstrap distribution for the mean is close to Normal, and the bootstrap confidence intervals (typical ranges on the right) are usually similar to the usual t interval, but slightly narrower. Bootstrapping the median produces a clearly non-Normal distribution; the bootstrap t interval should not be used for the median. (Ranges for median intervals are not given.) **(b)** The ratio of means is 1.0656; the bootstrap distribution is noticeably skewed, so the bootstrap t is not a good choice, but the other methods usually give intervals similar to 0.75 to 1.55. Also shown below is the bootstrap distribution for the ratio of the medians. It is considerably less erratic than the median ratio, but we have still not included these confidence intervals. **(c)** For example, the usual t interval from part (a) could be summarized by the statement, "On average, Jocko's estimates are 1% to 11% higher than those from other garages."

Typical ranges	
(a) Mean ratio	
t lower	1.00 to 1.02
t upper	1.10 to 1.12
Percentile lower	1.00 to 1.03
Percentile upper	1.09 to 1.11
BCa lower	1.00 to 1.03
BCa upper	1.09 to 1.11
(b) Ratio of means	
t lower	0.59 to 0.68
t upper	1.46 to 1.54
Percentile lower	0.69 to 0.78
Percentile upper	1.45 to 1.64
BCa lower	0.69 to 0.78
BCa upper	1.45 to 1.66





Normal quantile plot: median(Jocko / Other)
median(Jocko / Other)
Normal score
Bootstrap distribution: median(Jocko / Other)
Observed
Mean
boot t
percentile
BCa
median(Jocko / Other)
Normal quantile plot: mean(Jocko) / mean(Other)
mean(Jocko) / mean(Other)
Normal score
Bootstrap distribution: mean(Jocko) / mean(Other)
Observed
Mean
boot t
percentile
BCa
mean(Jocko) / mean(Other)
Normal quantile plot: median(Jocko) / median(Other)
median(Jocko) / median(Other)
Normal score
Bootstrap distribution: median(Jocko) / median(Other)
Observed
Mean
boot t
percentile
BCa
median(Jocko) / median(Other)

Chapter 17 Solutions

17.4. Possible examples of special causes might include: traffic, number of passengers on the shuttle (especially if the shuttle makes several stops along the way), mechanical problems with the shuttle.

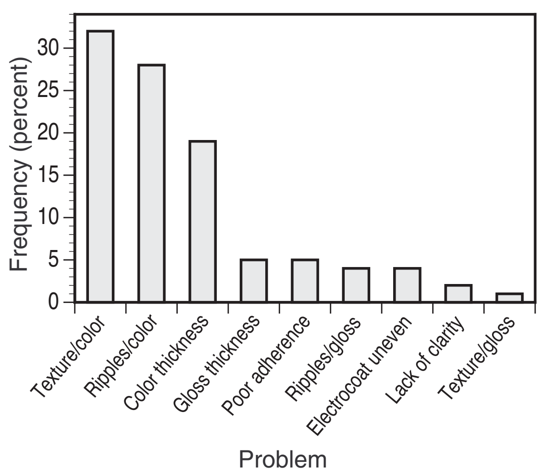
17.5. The center line is at $\mu = 85$ seconds. The control limits should be at $\mu \pm 3\sigma/\sqrt{6} = 85 \pm 3(17/\sqrt{6})$, which means about 64.18 and 105.82 seconds.

17.6. **(a)** With $n = 5$, the center line is unchanged (85 seconds), but the control limits are now $\mu \pm 3\sigma/\sqrt{5} = 62.18$ and 107.82 seconds. **(b)** With $n = 7$, the center line is unchanged (85 seconds), but the control limits are now $\mu \pm 3\sigma/\sqrt{7} \doteq 65.72$ and 104.28 seconds. **(c)** To convert to minutes, divide the original center line and control limits by 60: The center line is 1.417 minutes, and the control limits are 1.070 and 1.764 minutes.

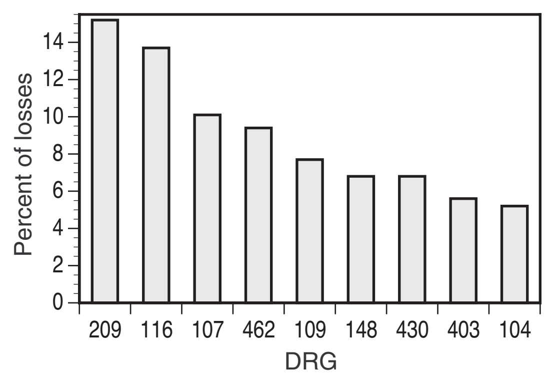
17.8. Common causes of variation might include the time it takes to call in the order, to make the pizza, and to deliver it. Examples of special causes might include heavy traffic or waiting for a train (causing delays in delivery), high demand for pizza (for example, during events like the Super Bowl), etc.

17.9. The most common problems are related to the application of the color coat; that should be the focus of our initial efforts.

For 17.9



For 17.10

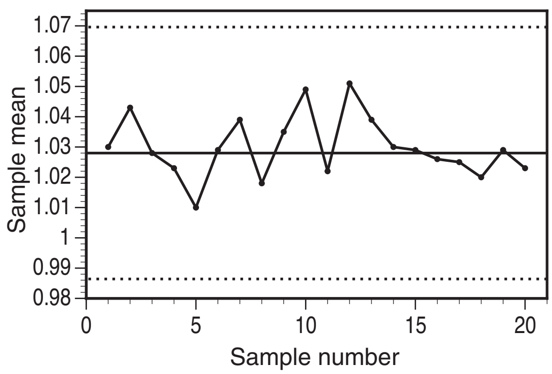


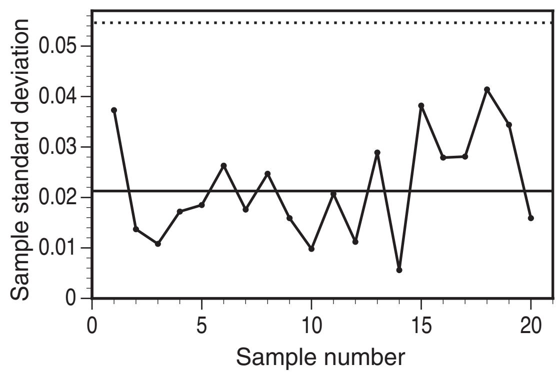
17.10. These DRGs account for a total of 80.5% of all losses. Certainly the first two (209 and 116) should be among those that are studied first; some students may also include 107, 462, and so on.

17.11. Possible causes could include delivery delays due to traffic or a train, high demand during special events, and so forth.

17.12. **(a)** The center line is at $\mu = 72^\circ$ F; the control limits should be at $\mu \pm 3\sigma/\sqrt{5}$, which means about 71.46° F and 72.54° F. **(b)** For $n = 5$, $c_4 = 0.94$ and $B_6 = 1.964$, so the center line for the s chart is $(0.94)(0.4) = 0.376^\circ$ F, and the control limits are 0 and 0.7856° F.

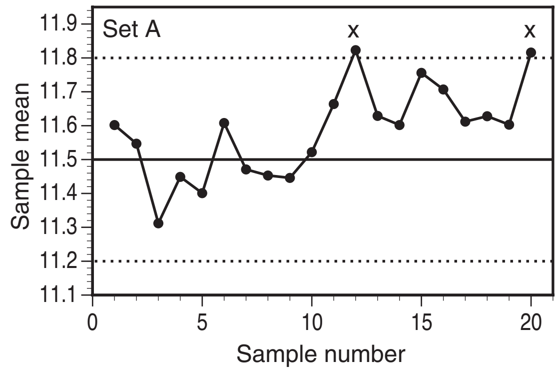
17.13. **(a)** For the $\bar{x}$ chart, the center line is at $\mu = 1.028$ lb; the control limits should be at $\mu \pm 3\sigma/\sqrt{3}$, which means about 0.9864 and 1.0696 lb. **(b)** For $n = 3$, $c_4 = 0.8862$ and $B_6 = 2.276$, so the center line for the s chart is $(0.8862)(0.024) = 0.02127$ lb, and the control limits are 0 and 0.05462 lb. **(c)** The control charts are below. **(d)** Both charts suggest that the process is in control.

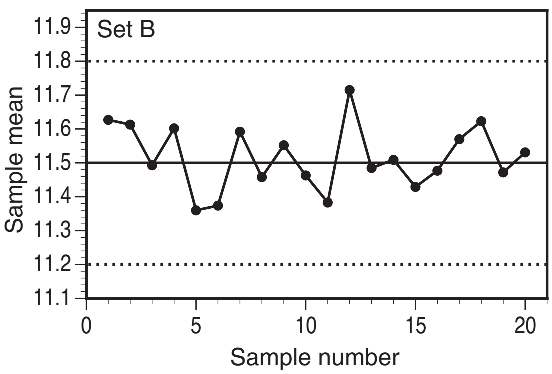


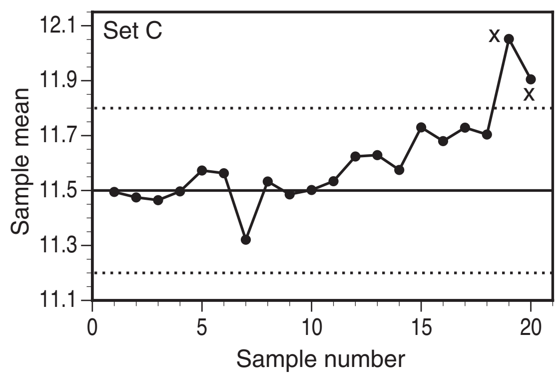


17.14. **(a)** Common causes might include processing time, normal workload fluctuation, or postal delivery time. **(b)** s-type special causes might include a new employee working in the personnel department. **(c)** Special causes affecting $\bar{x}$ might include a sudden large influx of applications or perhaps introducing a new filing system for applications.

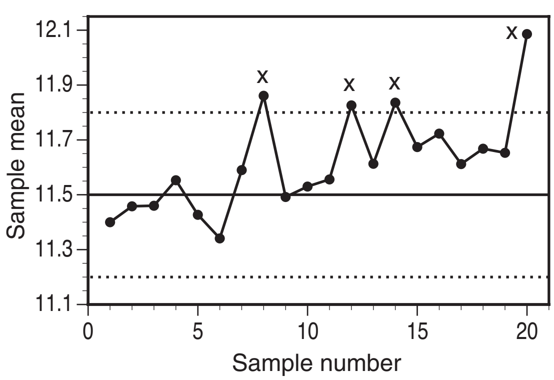
17.15. (a) The center line is at $\mu = 11.5$ Kp; the control limits should be at $\mu \pm 3\sigma/\sqrt{4} = 11.5 \pm 0.3 = 11.2$ and 11.8 Kp. **(b)** Graphs on the right and below. Points outside control limits are marked with an "X." **(c)** Set B is from the in-control process. The process mean shifted suddenly for Set A; it appears to have changed on about the 11th or 12th sample. The mean drifted gradually for the process in Set C.

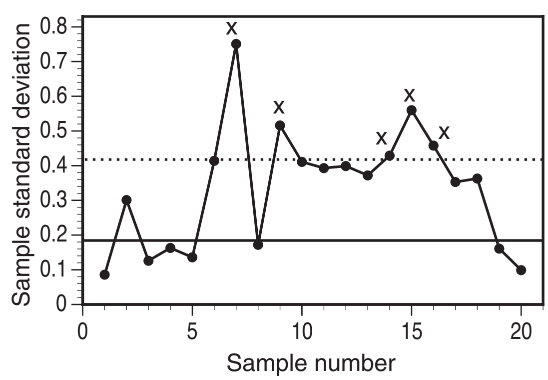






17.16. (a) For the $\bar{x}$ chart, the center line is 11.5, and the control limits are 11.2 and 11.8 (as in Exercise 17.15). For $n = 4$, $c_4 = 0.9213$ and $B_6 = 2.088$, so the center line for the s chart is $(0.9213)(0.2) = 0.18426$, and the control limits are 0 and 0.4176. **(b)** The s chart is certainly out of control at sample 7 (and was barely in control at sample 6). After that, there are a number of out-of-control points. The $\bar{x}$ chart is noticeably out of control at sample 8. **(c)** A change in the mean does not affect the s chart; the effect on the $\bar{x}$ chart is masked by the change in σ: Because of the increased variability, the sample means are sometimes below the UCL even after the process mean shifts.

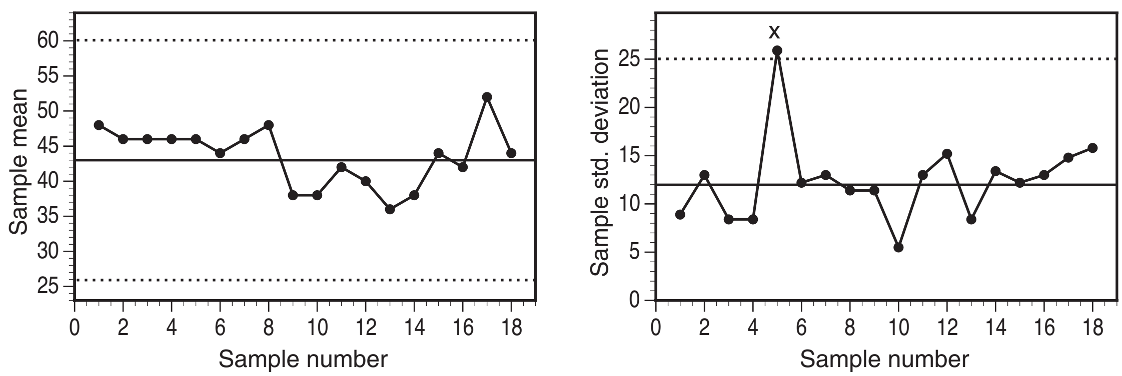




17.17. For the s chart with $n = 6$, we have $c_4 = 0.9515$, $B_5 = 0.029$ and $B_6 = 1.874$, so the center line is $(0.9515)(0.001) = 0.0009515$ inch, and the control limits are 0.000029 and 0.001874 inch. For the $\bar{x}$ chart, the center line is $\mu = 0.87$ inch, and the control limits are $\mu \pm 3\sigma/\sqrt{6} \doteq 0.87 \pm 0.00122 \doteq 0.8688$ and 0.8712 inch.

17.18. **(a)** For $n = 5$, we have $c_4 = 0.94$, $B_5 = 0$, and $B_6 = 1.964$, so the center line is 0.11938, and the control limits are 0 and 0.249428. **(b)** The center line is $\mu = 4.22$, and the control limits are $\mu \pm 3\sigma/\sqrt{5} \doteq 4.00496$ to 4.3904.

17.19. For the $\bar{x}$ chart, the center line is 43, and the control limits are 25.91 and 60.09. For $n = 5$, $c_4 = 0.9400$ and $B_6 = 1.964$, so the center line for the s chart is $(0.9400)(12.74) = 11.9756$, and the control limits are 0 and 25.02. The control charts (below) show that sample 5 was above the UCL on the s chart, but it appears to have been special cause variation, as there is no indication that the samples that followed it were out of control.



17.20. The new type of yarn would appear on the $\bar{x}$ chart because it would cause a shift in the mean pH. (It might also affect the process variability and therefore show up on the s chart.) Additional water in the kettle would change the pH for that kettle, which would change the mean pH and also change the process variability, so we would expect that special cause to show up on both the $\bar{x}$ and s charts.

17.21. **(a)** The process mean is the same as the center line: $\mu = 715$. The control limits are three standard errors from the mean, so $30 = 3\sigma/\sqrt{4}$, meaning that $\sigma = 20$. **(b)** If the mean changes to $\mu = 700$, then $\bar{x}$ is approximately Normal with mean 700 and standard deviation $\sigma/\sqrt{4} = 10$, so $\bar{x}$ will fall outside the control limits with probability $1 - P(685 < \bar{x} < 745) = 1 - P(-1.5 < Z < 4.5) = 0.0668$. **(c)** With $\mu = 700$ and $\sigma = 30$, $\bar{x}$ is approximately Normal with mean 700 and standard deviation $\sigma/\sqrt{4} = 15$, so $\bar{x}$ will fall outside the control limits with probability $1 - P(685 < \bar{x} < 745) = 1 - P(-1 < Z < 3) = 0.16$.

17.22. $c = 3.090$. (Looking at Table A, there appear to be three possible answers—3.08, 3.09, or 3.10. Software gives the answer 3.090232. . . .)

17.23. The usual 3σ limits are $\mu \pm 3\sigma/\sqrt{n}$ for an $\bar{x}$ chart and $(c_4 \pm 3c_5)\sigma$ for an s chart. For 2σ limits, simply replace "3" with "2." **(a)** $\mu \pm 2\sigma/\sqrt{n}$. **(b)** $(c_4 \pm 2c_5)\sigma$.

17.24. **(a)** The R chart monitors the *variability* (spread) of the process. **(b)** The R chart is commonly used because R is easier to compute (by hand) than s. **(c)** The $\bar{x}$ control limits are affected because we estimate process spread using R instead of s.

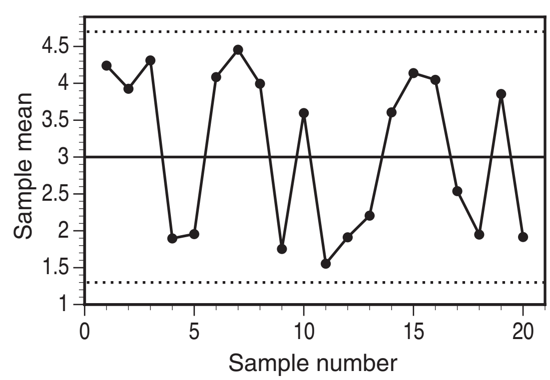
17.25. **(a)** Shrinking the control limits would increase the frequency of false alarms, because the probability of an out-of-control point when the process is in control will be higher (roughly 5% instead of 0.3%). **(b)** Quicker response comes at the cost of more false alarms. **(c)** The runs rule is better at detecting gradual changes. (The one-point-out rule is generally better for sudden, large changes.)

17.26. **(a)** Either (ii) or (iii), depending on whether the deterioration happens quickly or gradually. We would not necessarily expect that this deterioration would result in a change in variability (s or R). **(b)** (i) s or R chart: A change in precision suggests altered variability (s or R), but not necessarily a change in center ($\bar{x}$). **(c)** (i) s or R chart: Assuming there are other (fluent) customer service representatives answering the phones, this new person would have unusually long times, which should most quickly show up as an increase in variability. **(d)** (iii) A run on the $\bar{x}$ chart: "The runs signal responds to a gradual shift more quickly than the one-point-out signal."

17.27. We estimate $\hat{\sigma}$ to be $\bar{s}/0.9213 \doteq 1.1180$, so the $\bar{x}$ chart has center line $\bar{\bar{x}} = 47.2$ and control limits $\bar{\bar{x}} \pm 3\hat{\sigma}/\sqrt{4} \doteq 45.523$ and 48.877. The s chart has center line $\bar{s} = 1.03$ and control limits 0 and $2.088\hat{\sigma} \doteq 2.3344$.

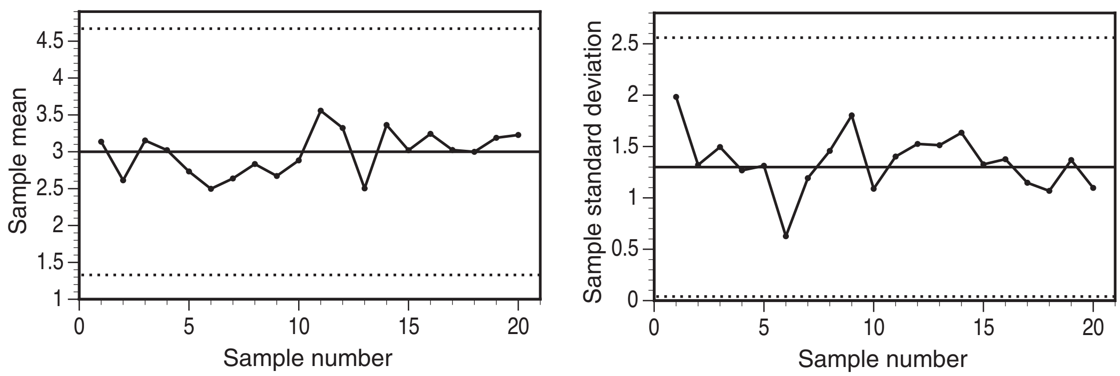
17.28. To estimate μ and σ, we compute $\bar{\bar{x}} \doteq 1.0299$ lb and $\bar{s} \doteq 0.0222$ lb from the sample means and standard deviations given in Table 17.3. $\hat{\mu} = \bar{\bar{x}}$ is our estimate of μ; this is about 0.002 lb greater than the historical value (1.028 lb). To estimate σ, we use $\hat{\sigma} = \bar{s}/c_4 = 0.0222/0.8862 \doteq 0.0251$ lb—about 4.6% greater than the historical value (0.024 lb). Both of these differences are so small that, even if they are statistically significant, it seems unlikely that they suggest any noteworthy change in this process.

17.29. One possible $\bar{x}$ chart is shown, created with the (arbitrary) assumption that the experienced clerk processes invoices in an average of 2 minutes, while the new hire takes an average of 4 minutes. (The control limits were set arbitrarily as well.)



Note: *Such a process would not be considered to be in control for very long. The initial control limits might be developed based on a historical estimate of σ, but eventually we should assess that estimate based on our sample standard deviations. Because both clerks "are quite consistent, so that their times vary little from invoice to invoice," each sample has a small value of s, so the revised estimate of σ would likely be smaller. At that point, the control limits (based on that smaller spread) will be moved closer to the center line.*

17.30. **(a)** Sketches will vary quite a bit; many students will struggle with the implications of this situation on the appearance of the two charts. The two charts below were produced using a much more sophisticated approach than most students would take; they arose from a simulation taking samples of size 6 (3 from each clerk), where the experienced clerk's processing time (in minutes) is $N(2, 0.5)$ and the new hire's processing time is $N(4, 0.8)$. The center lines and control limits were estimated from the data. **(b)** For example, this would be acceptable if we are concerned with *overall* processing time and are not interested in individual processing times. In particular, it would not be appropriate to compute tolerance limits for this situation because the individual measurements do not have a Normal distribution.



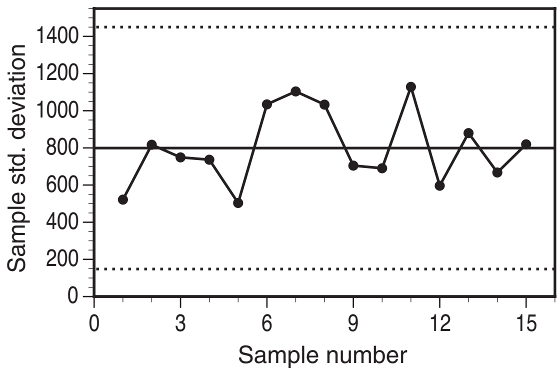
Note: *The situation described here demonstrates the problems that can arise when we do not carefully consider the question of "rational subgroups" in our sampling design; see the discussion on page 17-30.*

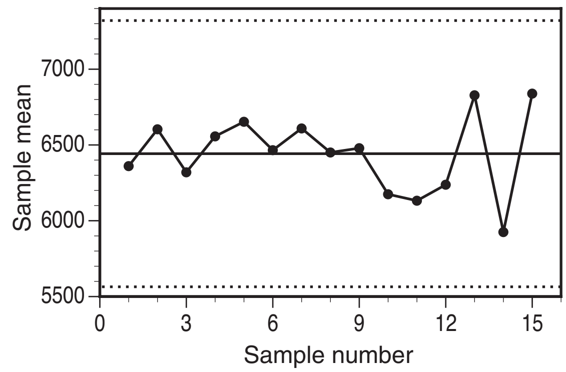
This situation is related to the issue of distinguishing "within-groups" variation from "between-groups" variation, as discussed in Chapter 12 (One-Way ANOVA). The within-groups variation is variation in invoice processing time for each clerk, and the between-groups variation is the difference between their processing times. In this case, though, we are not paying attention to the explanatory variable (which clerk processed the invoice), so all we see is a mixture of the two sources of variation. If—as was the case here—the two clerks were fairly consistent so that within-groups variation is small, the sample standard deviations are most affected by the between-groups variation.

Note that both charts show less variation than we typically see; nearly all the points are no more than 1 or 2 standard deviations from the center line. To begin to understand why, imagine an extreme case with no within-groups variation—where one clerk always takes exactly 2 minutes, and the other always takes exactly 4 minutes. Then each sample would contain the 6 numbers 2, 2, 2, 4, 4, and 4, so $\bar{x} = 3$ and $s \doteq 1.0954$ for all samples, and the control charts would have no variation at all.

17.31. **(a)** Average the 20 sample means and standard deviations and estimate μ to be $\hat{\mu} = \bar{\bar{x}} = 2750.7$ and σ to be $\hat{\sigma} = \bar{s}/c_4 = 345.5/0.9213 \doteq 375.0$. **(b)** In the s chart shown in Figure 17.7, most of the points fall below the center line.

17.32. For the 15 samples, we have $\bar{s} = \$799.1$ and $\bar{\bar{x}} = \$6442.4$.
(a) $\hat{\sigma} = \bar{s}/c_4 = 799.1/0.9650 = 828.1$; the center line is $\bar{s}$, and the control limits are $B_5\hat{\sigma} = (0.179)(\$828.1) = \$148.2$ and $B_6\hat{\sigma} = (1.751)(\$828.1) = \$1450.0$. **(b)** For the $\bar{x}$ chart, the center line is $\bar{\bar{x}} = \$6442.4$, and the control limits are $\bar{\bar{x}} \pm 3\hat{\sigma}/\sqrt{8} = \$5564.1$ to $\$7320.7$. The control chart shows that the process is in control.

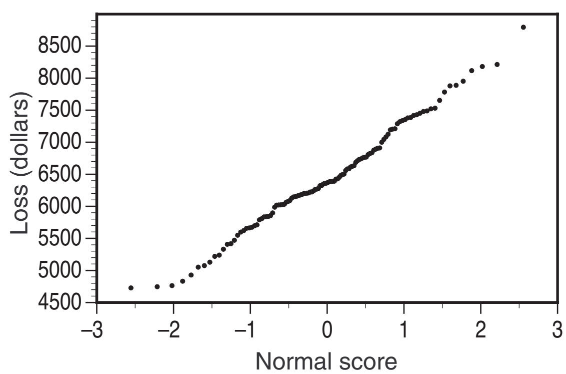




17.33. If the manufacturer practices SPC, that provides some assurance that the phones are roughly uniform in quality—as the text says, "We know what to expect in the finished product." So, assuming that uniform quality is sufficiently high, the purchaser does not need to inspect the phones as they arrive because SPC has already achieved the goal of that inspection: to avoid buying many faulty phones. (Of course, a few unacceptable phones may be produced and sold even when SPC is practiced—but inspection would not catch all such phones anyway.)

17.34. The standard deviation of all 120 measurements is $s \doteq \$811.53$, and the mean is $\bar{x} \doteq \$6442.4$ (the same as $\bar{\bar{x}}$—as it must be, provided all the individual samples were the same size). The natural tolerances are $\bar{x} \pm 3s = \$4007.8$ to $\$8877.0$.

17.35. The quantile plot does not suggest any serious deviations from Normality, so the natural tolerances should be reasonably trustworthy.



Note: *We might also assess Normality with a histogram or stemplot; this looks reasonably Normal, but we see that the number of losses between \$6000 and \$6500 is noticeably higher than we might expect from a Normal distribution. In fact, the smallest and largest losses were \$4727 and \$8794. These are both within the tolerances, but note that the minimum is quite a bit more than the lower limit of the tolerances (\$4008). The large number of losses between \$6000 and \$6500 makes the mean slightly lower and therefore lowers both of the tolerance limits.*

17.36. (a) About 99.9% meet the old specifications: If X is the water resistance on a randomly chosen jacket, then:

$$P(1000 < X < 4000) = P\left(\frac{1000-2750}{383.8} < Z < \frac{4000-2750}{383.8}\right) = P(-4.56 < Z < 3.26) \doteq 0.9994$$

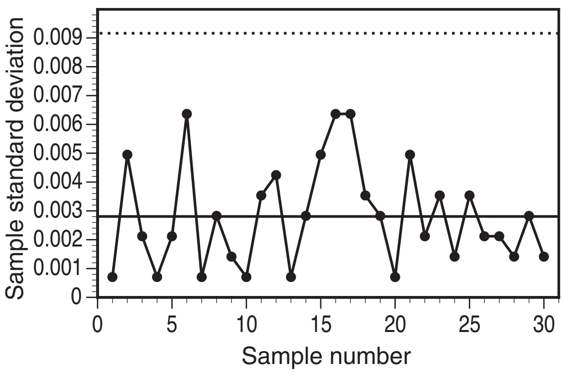
(b) About 97.4% meet the new specifications:

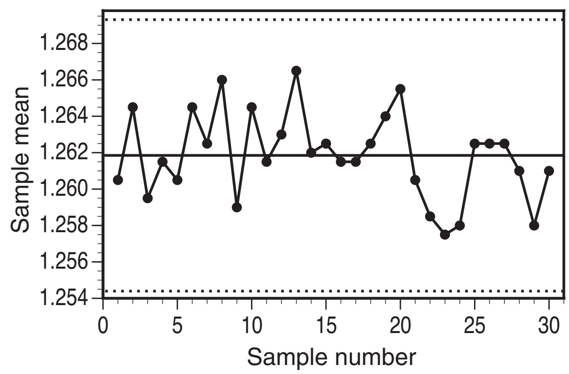
$$P(1500 < X < 3500) = P\left(\frac{1500-2750}{383.8} < Z < \frac{3500-2750}{383.8}\right) = P(-3.26 < Z < 1.95) \doteq 0.9738$$

17.37. If we shift the process mean to 2500 mm, about 99% will meet the new specifications:

$$P(1500 < X < 3500) = P\left(\frac{1500-2500}{383.8} < Z < \frac{3500-2500}{383.8}\right) = P(-2.61 < Z < 2.61) \doteq 0.9910$$

17.38. **(a)** The means (1.2605 and 1.2645) agree exactly with those given; the standard deviations are the same up to rounding. **(b)** The s chart tracks process spread. For the 30 samples, we have $\bar{s} = 0.0028048$, so $\hat{\sigma} = \bar{s}/c_4 = \bar{s}/0.7979 \doteq 0.003515$; the center line is $\bar{s}$, and the control limits are $B_5\hat{\sigma} = 0$ and $B_6\hat{\sigma} = 2.606\hat{\sigma} \doteq 0.009161$. Short-term variation seems to be in control. **(c)** For the $\bar{x}$ chart, which monitors the process center, the center line is $\bar{\bar{x}} = 1.26185$, and the control limits are $\bar{\bar{x}} \pm 3\hat{\sigma}/\sqrt{2} \doteq 1.2544$ to 1.2693. The control chart shows that the process is in control.





17.39. The mean of the 17 in-control samples is $\bar{\bar{x}} = 43.4118$, and the standard deviation is 11.5833, so the natural tolerances are $\bar{\bar{x}} \pm 3s = 8.66$ to 78.16.

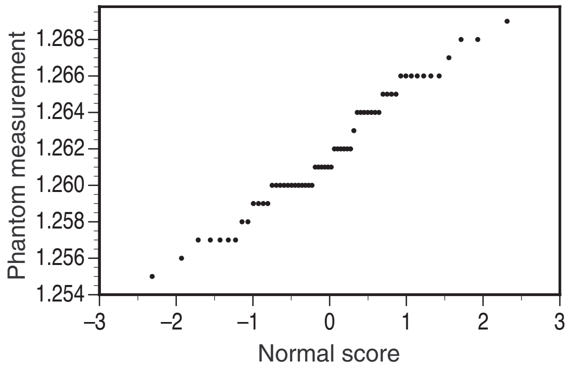
17.40. There were no out-of-control points, so we estimate the mean of the process using $\hat{\mu} = \bar{\bar{x}} = 1.26185$. The estimated standard deviation is computed from the 60 individual data points; this gives $s \doteq 0.003328$. The natural tolerances are $\bar{\bar{x}} \pm 3s = 1.2519$ to 1.2718.

17.41. Only about 44% of meters meet the specifications. Using the mean (43.4118) and standard deviation (11.5833) found in the solution to Exercise 17.39:

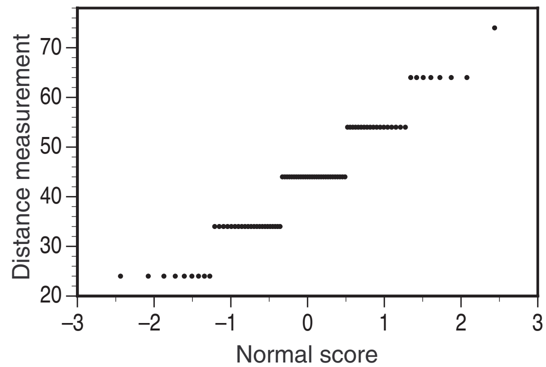
$$P(44 < X < 64) = P\left(\frac{44-43.4118}{11.5833} < Z < \frac{64-43.4118}{11.5833}\right) = P(0.05 < Z < 1.78) \doteq 0.4426$$

17.42. There is no clear deviation from Normality apart from granularity due to the limited accuracy of the recorded measurements.

For 17.42



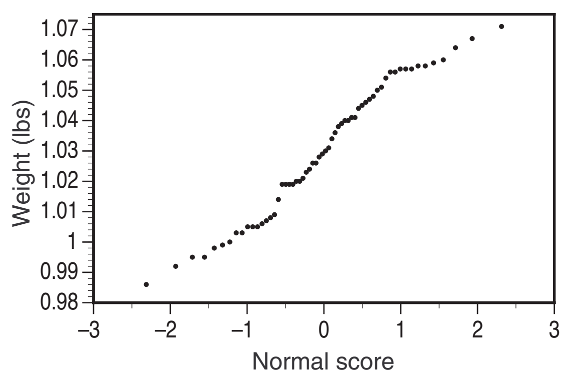
For 17.43



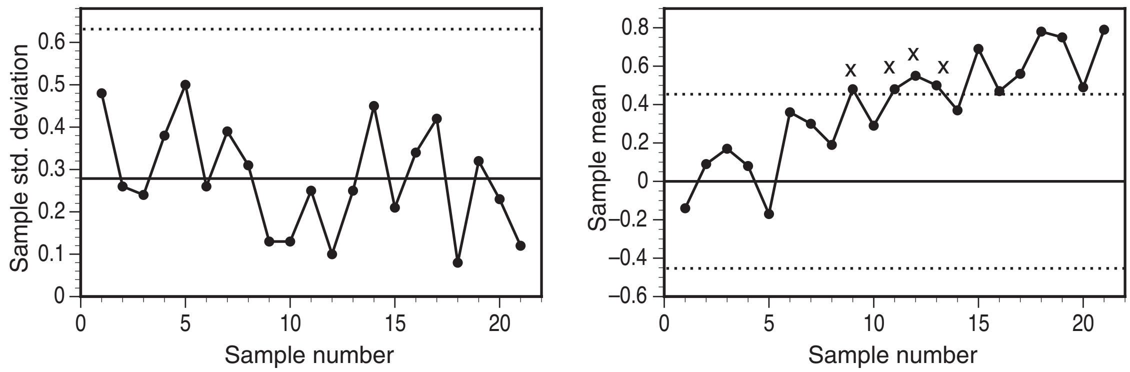
17.43. The limited precision of the measurements shows up in the granularity (stair-step appearance) of the graph. Aside from this, there is no particular departure from Normality.

17.44. The standard deviation of all 60 weights is $s \doteq 0.0224$ lb, and the mean is $\bar{x} \doteq 1.0299\overline{6}$ lb (the same as $\bar{\bar{x}}$, except for rounding error). The natural tolerances are $\bar{x} \pm 3s = 0.9627$ to 1.0972 lb.

17.45. The quantile plot, while not perfectly linear, does not suggest any serious deviations from Normality, so the natural tolerances should be reasonably trustworthy.



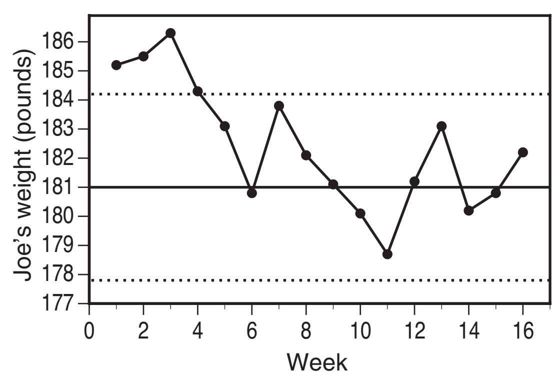
17.46. (a) For the 21 samples, we have $\bar{s} \doteq 0.2786$, so $\hat{\sigma} = \bar{s}/c_4 = 0.2786/0.9213 \doteq 0.3024$; the center line is $\bar{s}$, and the control limits are $B_5\hat{\sigma} = 0$ and $B_6\hat{\sigma} = (2.088)(0.3024) \doteq 0.6313$. Short-term variation seems to be in control. **(b)** For the $\bar{x}$ chart, the center line is 0 and the control limits are $\pm 3\hat{\sigma}/\sqrt{4} = \pm 0.4536$. The $\bar{x}$ chart suggests that the process mean has drifted. (Only the first four out-of-control points are marked.) One possible cause for the increase in the mean is that the machine that makes the bearings is gradually drifting out of adjustment.



17.47. **(a)** (ii) A sudden change in the $\bar{x}$ chart: This would immediately increase the amount of time required to complete the checks. **(b)** (i) A sudden change (decrease) in s or R because the new measurement system will remove (or decrease) the variability introduced by human error. **(c)** (iii) A gradual drift in the $\bar{x}$ chart (presumably a drift up, if the variable being tracked is the length of time to complete a set of invoices).

17.49. The process is no longer the same as it was during the downward trend (from the 1950s into the 1980s). In particular, including those years in the data used to establish the control limits results in a mean that is too high to use for current winning times, and a standard deviation that includes variation attributable to the "special cause" of the changing conditioning and professional status of the best runners. Such special cause variation should not be included in a control chart.

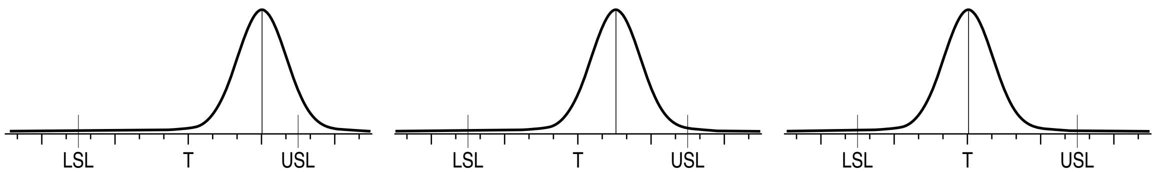
17.50. The center line is 181 pounds and the control limits are $181 \pm 3.2 = 177.8$ and 184.2 pounds. The first four points are above the upper control limit; there are no runs (above or below the center line) longer than five. The overall impression is that Joe's weight returns to being "in control"; it decreases fairly steadily, and the last 12 points are between the control limits.



17.51. LSL and USL are specification limits on the individual observations. This means that they do not apply to averages and that they are *specified* as desired output levels, rather than being *computed* based on observation of the process. LCL and UCL are control limits for the averages of samples drawn from the process. They may be determined from past data, or independently specified, but the main distinction is that the purpose of control limits is to detect whether the process is functioning "as usual," while specification limits are used to determine what percentage of the outputs meet certain specifications (are acceptable for use).

17.52. In each graph below, the large tick marks are 3σ apart, and the smaller tick marks are 1σ apart, and the target is marked as "T." Because $C_p = 1.5$, the specification limits are 9σ apart, located at $T \pm 4.5\sigma$. The first two graphs could be flipped (i.e., the peak of the curve could be closer to the LSL than the USL). **(a)** $C_{pk} = 0.5$ means that the nearer specification limit is $(0.5)(3\sigma) = 1.5\sigma$ above (or below) the mean. **(b)** $C_{pk} = 1.0$ means that the nearer specification limit is 3σ above (or below) the mean. **(c)** $C_{pk} = 1.5$ means that the nearer specification limit is $(1.5)(3\sigma) = 4.5\sigma$ above (or below) the mean—so that μ falls exactly on the target (halfway between the specification limits).

Note: *At the end of Example 17.16, the text notes that $C_p = C_{pk}$ means "the process is properly centered"—that is, μ equals the target.*



17.53. For computing $\hat{C}_{pk}$, note that the estimated process mean (2750.7 mm) lies closer to the USL. **(a)** $\hat{C}_p = \dfrac{4000 - 1000}{6 \times 383.8} \doteq 1.3028$ and $\hat{C}_{pk} = \dfrac{4000 - 2750.7}{3 \times 383.8} \doteq 1.0850$. **(b)** $\hat{C}_p = \dfrac{3500 - 1500}{6 \times 383.8} \doteq 0.8685$ and $\hat{C}_{pk} = \dfrac{3500 - 2750.7}{3 \times 383.8} \doteq 0.6508$.

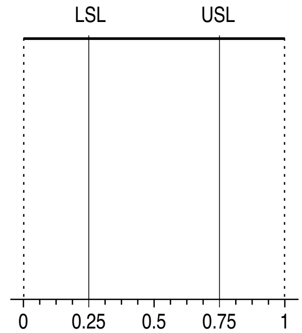
17.54. (a) With the original specifications, $\hat{C}_p \doteq 1.3028$ (unchanged from the previous exercise, because $\hat{C}_p$ does not depend on μ) and $\hat{C}_{pk} = \dfrac{4000 - 2500}{3 \times 383.8} \doteq 1.3028$. **(b)** Once again, $\hat{C}_p \doteq 0.8685$ is unchanged. $\hat{C}_{pk} = \dfrac{3500 - 2500}{3 \times 383.8} \doteq 0.8685$.

17.55. In the solution to Exercise 17.44, we found that the mean and standard deviation of all 60 weights are $\bar{x} \doteq 1.0299\overline{6}$ lb and $s \doteq 0.0224$ lb. **(a)** $\hat{C}_p = \dfrac{1.10 - 0.94}{6 \times 0.0224} \doteq 1.1901$ and $\hat{C}_{pk} = \dfrac{1.10 - 1.03}{3 \times 0.0224} \doteq 1.0418$. (These were computed with the unrounded values of $\bar{x}$ and s; rounding will produce slightly different results.) **(b)** Customers typically will not complain about a package that was too heavy.

17.56. A change to the process mean would not change $\hat{C}_p$, but we could increase $\hat{C}_{pk}$ by centering the process mean between the specification limits, at $\mu = \frac{1.10+0.94}{2} = 1.02$ lb. With that change, $\hat{C}_{pk}$ increases to 1.1901 (the same as $\hat{C}_p$ before the change).

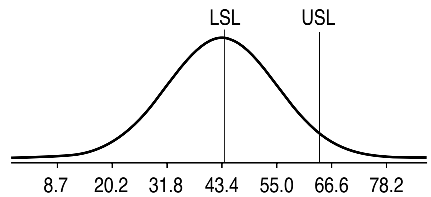
Note: *The effect of this change is hard to predict if we suspect that the weight measurements are non-Normal, but the data do not suggest any such problems (see the solution to Exercise 17.45). Additionally, decreasing the process mean might have the undesirable effect of increasing customer dissatisfaction (see part (b) of the previous exercise).*

17.57. **(a)** $C_{pk} = \frac{0.75-0.25}{3\sigma} \doteq 0.5767$. 50% of the output meets the specifications. **(b)** LSL and USL are 0.865 standard deviations above and below to mean, so the proportion meeting specifications is $P(-0.865 < Z < 0.865) \doteq 0.6130$. **(c)** The relationship between C_{pk} and the proportion of the output meeting specifications depends on the shape of the distribution.



17.58. In the solution to Exercise 17.31, we found $\hat{\sigma} = \bar{s}/c_4 \doteq 375.0$; from this, we compute $\hat{C}_{pk} = \dfrac{3500-2500}{3 \times 375.0} \doteq 0.8889$, which is larger than the previous value (0.8685).

17.59. See also the solution to Exercise 17.43. **(a)** Use the mean and standard deviation of the 85 remaining observations: $\hat{\mu} = \bar{x} = 43.4118$ and $\hat{\sigma} = s = 11.5833$. **(b)** $\hat{C}_p = \frac{20}{6\hat{\sigma}} \doteq 0.2878$ and $\hat{C}_{pk} = 0$ (because $\hat{\mu}$ is outside the specification limits). This process has very poor capability: The mean is too low and the spread too great. Only about 46% of the process output meets specifications.



17.60. See also the solution to Exercise 17.34. **(a)** About 97.1%: For the 120 observations in Table 17.7, we find $\hat{\mu} = \bar{x} \doteq \$6442.4$ and $\hat{\sigma} = s \doteq \$811.53$. Therefore, we estimate $P(\$4500 < X < \$7500) = P(-2.39 < Z < 1.30) = 0.9032 - 0.0084 = 0.8948$. **(b)** $\hat{C}_p = \dfrac{7500-4500}{6 \times 811.53} \doteq 0.6161$. **(c)** $\hat{C}_{pk} = \dfrac{7500-6442.4}{3 \times 811.53} \doteq 0.4344$.

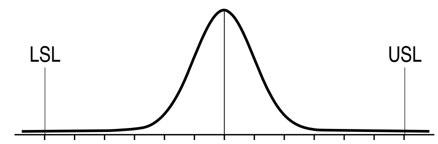
17.61. We have $\bar{x} = 22.005$ mm and $s = 0.009$ mm, so we assume that an individual bearing diameter X follows a $N(22.005, 0.009)$ distribution. **(a)** About 85.3% meet specifications:

$$\begin{aligned} P(21.985 < X < 22.015) &= P\left(\tfrac{21.985-22.005}{0.009} < Z < \tfrac{22.015-22.005}{0.009}\right) \\ &= P(-2.22 < Z < 1.11) \\ &= 0.9868 - 0.1335 = 0.8533. \end{aligned}$$

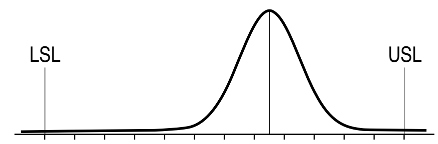
(b) $\hat{C}_{pk} = \dfrac{22.015-22.005}{3 \times 0.009} \doteq 0.3704$.

17.62. **(a)** This is unlikely to have any beneficial effect; it would result in more frequent adjustments, but these would often be unnecessary and so might degrade capability. Control limits are for correcting special-cause variation, not common-cause variation. **(b)** If some of the nonconforming bearings are due to operator error, further training may have the effect of reducing σ and increasing C_{pk}. Part (d) offers a slightly different viewpoint. **(c)** Assuming the new machine has less variability (smaller σ), this should improve the process capability. **(d)** The number of nonconforming bearings produced by an operator is (for the most part) a result of random variation within the system; no incentive can cause the operator to do better than the system allows. **(e)** Better raw material should (presumably) result in better product, so this should improve the capability.

17.63. This graph shows a process with Normal output and $C_p = 2$. The tick marks are σ units apart; this is called "six-sigma quality" because the specification limits are (at least) six standard deviations above and below the mean.

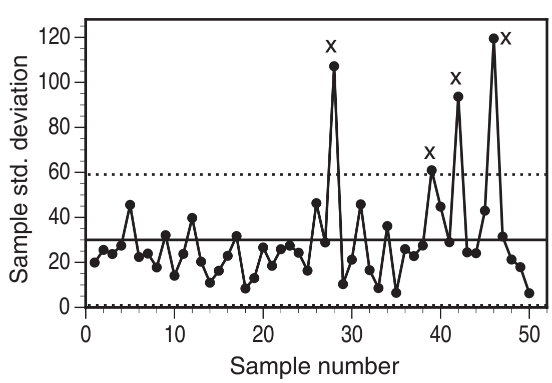


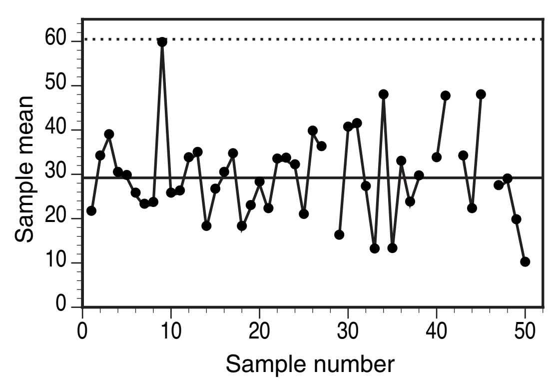
17.64. **(a)** The graph on the right shows the mean shifted toward the USL; it could also be shifted toward the LSL. As in the graph in the previous problem, tick marks are σ units apart. **(b)** $C_{pk} = \frac{4.5\sigma}{3\sigma} = 1.5$. Six-sigma quality does *not* mean that $C_{pk} \geq 2$; the latter is a stronger requirement. **(c)** The desired probability is $1 - P(-7.5 < Z < 4.5)$, for which software gives 3.4×10^{-6}, or about 3.4 out-of-spec parts per million.



17.65. Students will have varying justifications for the sampling choice. Choosing six calls per shift gives an idea of the variability and mean for the shift as a whole. If we took six consecutive calls (at a randomly chosen time), we might see additional variability in $\bar{x}$ because sometimes those six calls might be observed at particularly busy times (when a customer has to wait for a long time until a representative is available or when a representative is using the restroom).

17.66. **(a)** For $n = 6$, we have $c_4 = 0.9515$, $B_5 = 0.029$, and $B_6 = 1.874$. With $\bar{s} = 29.985$ seconds, we compute $\hat{\sigma} = \bar{s}/c_4 \doteq 31.5134$ seconds, so the initial s chart has center line $\bar{s}$ and control limits $B_5\hat{\sigma} \doteq 0.9139$ and $B_6\hat{\sigma} \doteq 59.0561$ seconds. There are four out-of-control points, from samples 28, 39, 42, and 46. **(b)** With the remaining 46 samples, $\bar{s} = 24.3015$, so $\hat{\sigma} = \bar{s}/c_4 = 25.54$ seconds, and the control limits are $B_5\hat{\sigma} = 0.741$ and $B_6\hat{\sigma} = 47.86$ seconds. There are no more out-of-control points. (The second s chart is not shown.) **(c)** We have center line $\bar{\bar{x}} = 29.2087$ seconds, and control limits $\bar{\bar{x}} \pm 3\hat{\sigma}/\sqrt{6} = -2.072$ and 60.489 seconds. (The lower control limit should be ignored or changed to 0.) The $\bar{x}$ chart has no out-of-control points.





17.67. The outliers are 276 seconds (sample 28), 244 seconds (sample 42), and 333 seconds (sample 46). After dropping those outliers, the standard deviations drop to 9.284, 6.708, and 31.011 seconds. (Sample #39, the other out-of-control point, has two moderately large times, 144 and 109 seconds; if they are removed, s drops to 3.416.)

17.68. For those 10 days, there were 961 absences and $10 \cdot 987 = 9870$ person-days available for work, so $\bar{p} = \frac{961}{9870} \doteq 0.09737$, and:

$$\text{CL} = \bar{p} = 0.09737, \text{control limits: } \bar{p} \pm 3\sqrt{\frac{\bar{p}(1-\bar{p})}{987}} = 0.06906 \text{ and } 0.12567$$

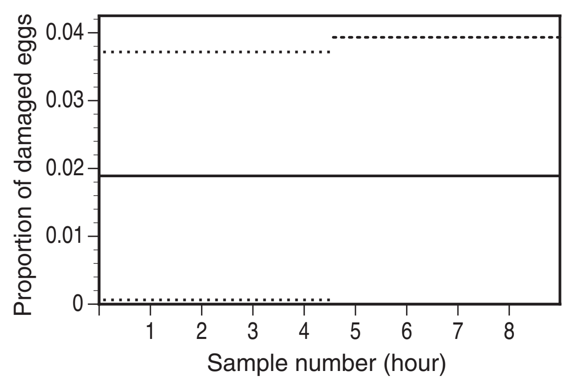
17.69. (a) For those 10 months, there were 1028 overdue invoices out of 28,400 total invoices (opportunities), so $\bar{p} = \frac{1028}{28,400} \doteq 0.03620$. **(b)** The center line and control limits are:

$$\text{CL} = \bar{p} = 0.03620, \text{control limits: } \bar{p} \pm 3\sqrt{\frac{\bar{p}(1-\bar{p})}{2840}} = 0.02568 \text{ and } 0.04671$$

17.70. Based on 3.22 complaints per 1000 passengers, the center line is $\bar{p} = \frac{3.22}{1000} = 0.00322$, and the control limits are $\bar{p} \pm 3\sqrt{\frac{\bar{p}(1-\bar{p})}{2500}}$, which means about -0.000179 and 0.006619. As the problem says, we take LCL $= 0$.

17.71. The center line is at the historical rate (0.0189); the control limits are $0.0189 \pm 3\sqrt{\frac{0.0189 \cdot 0.9811}{500}}$, which means about 0.00063 and 0.03717.

17.72. For both operators, the center line is 0.0189. For the first operator, the control limits are those found in the previous solution: 0.00063 and 0.03717. For the second operator, the control limits are $0.0189 \pm 3\sqrt{\frac{0.0189 \cdot 0.9811}{400}}$, which yields -0.00153 (use 0) and 0.03933.



Note: *We could simplify this control chart with a couple of practical observations. First, we would not be concerned if the proportion of broken eggs were too low, so we could take the first operator's LCL to be 0. In addition, for the first (second) operator, the hourly proportion will be a multiple of 0.002 (0.0025), so it will exceed the UCL if $\hat{p} \geq 0.038$ ($\hat{p} \geq 0.04$).*

17.73. The center line is at $\bar{p} = \frac{163}{36,480} \doteq 0.004468$; the control limits should be at $\bar{p} \pm 3\sqrt{\frac{\bar{p}(1-\bar{p})}{1520}}$, which means about -0.00066 (use 0) and 0.0096.

17.74. The initial center line and control limits are:

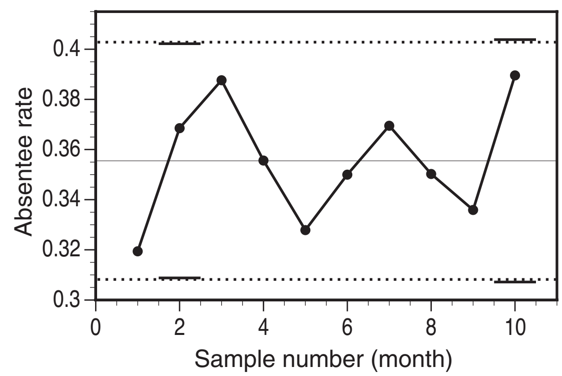
$$\text{CL} = p = 0.01, \text{control limits: } p \pm 3\sqrt{\frac{p(1-p)}{90,000}} = 0.009005 \text{ and } 0.010995$$

On a day when only 45,000 prescriptions are filled, the center line is unchanged, while the control limits change to:

$$p \pm 3\sqrt{\frac{p(1-p)}{45,000}} = 0.008593 \text{ and } 0.011407$$

17.75. **(a)** The student counts sum to 9218, while the absentee total is 3277, so $\bar{p} = \frac{3277}{9218} = 0.3555$ and $\bar{n} = 921.8$. **(b)** The center line is $\bar{p} = 0.3555$, and the control limits are:

$$\bar{p} \pm 3\sqrt{\frac{\bar{p}(1-\bar{p})}{921.8}} = 0.3082 \text{ and } 0.4028$$



The p chart suggests that absentee rates are in control. **(c)** For October, the limits are 0.3088 and 0.4022; for June, they are 0.3072 and 0.4038. These limits appear as solid lines on the p chart, but they are not substantially different from the control limits found in (b). Unless n varies *a lot* from sample to sample, it is sufficient to use $\bar{n}$.

17.76. **(a)** $\bar{p} = \frac{3.5}{1{,}000{,}000} = 0.0000035$. At 5000 pieces per day, we expect 0.0175 defects per day; in a 24-day month, we would expect 0.42 defects. **(b)** The center line is 0.0000035; assuming that every day we examine all 5000 pieces, the LCL is negative (so we use 0), and the UCL is 0.0000083. **(c)** Note that most of the time, we will find 0 defects, so that $\hat{p} = 0$. If we should ever find even one defect, we would have $\hat{p} = 0.0002$, and the process would be out of control. On top of this, it takes an absurd amount of testing in order to catch the rare defect.

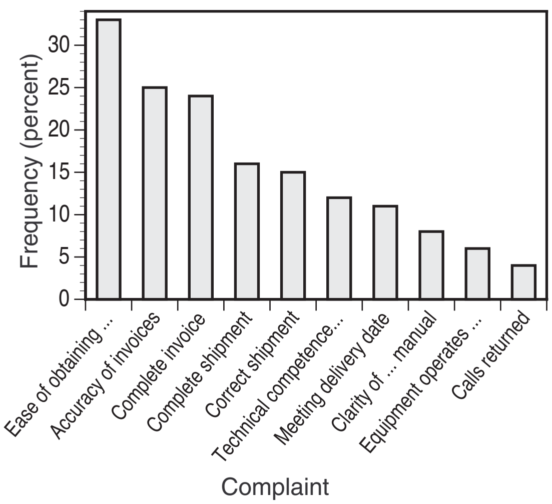
17.77. **(a)** $\bar{p} = \dfrac{8000}{1,000,000} = 0.008$. We expect about $4 = (500)(0.008)$ defective orders per month. **(b)** The center line and control limits are:

$$\text{CL} = \bar{p} = 0.008, \text{ control limits: } \bar{p} \pm 3\sqrt{\frac{\bar{p}(1-\bar{p})}{500}} = -0.00395 \text{ and } 0.01995$$

(We take the lower control limit to be 0.) It takes at least ten bad orders in a month to be out of control because $(500)(0.01995) = 9.975$.

17.78. Control charts focus on ensuring that the *process* is consistent, not that the *product* is good. An in-control process may consistently produce some percentage of low-quality products. Keeping a process in control allows one to detect shifts in the distribution of the output (which may have been caused by some correctable error); it does not help in fixing problems that are inherent to the process.

17.79. **(a)** The percents do not add to 100% because one customer might have several complaints; that is, he or she could be counted in several categories. **(b)** Clearly, top priority should be given to the process of creating, correcting, and adjusting invoices, as the three most common complaints involved invoices.

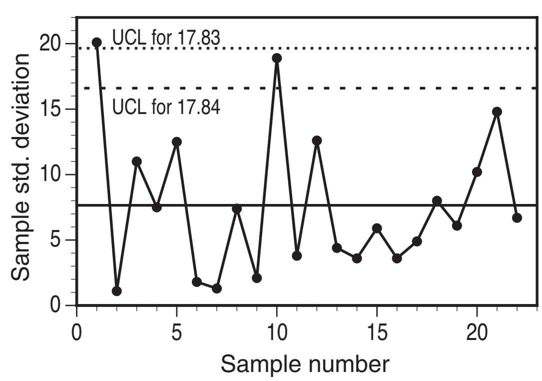


17.80. **(a)** Use $\bar{x}$ and s charts to track the time required. **(b)** Use a p chart to track the acceptance percentage. **(c)** Use $\bar{x}$ and s charts to track the thickness. **(d)** Use a p chart to track the proportion of dropped calls.

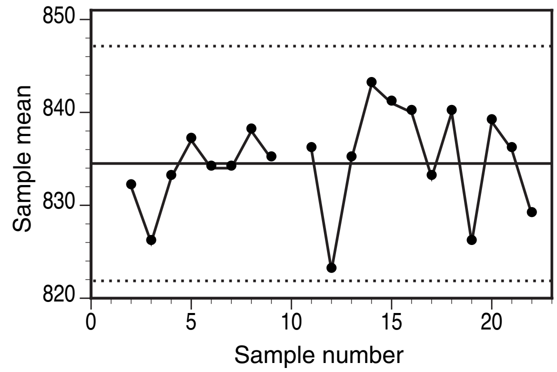
17.81. On one level, these two events are similar: Points below the LCL on an $\bar{x}$ (s) chart suggest that the process mean (standard deviation) may have decreased. The difference is in the implications of such a decrease (if not due to a special cause). For the mean, a decrease might signal a need to recalibrate the process in order to keep meeting specifications (that is, to bring the process back into control). A decrease in the standard deviation, on the other hand, typically does not indicate that adjustment or recalibration is necessary, but it will require re-computation of the $\bar{x}$ chart control limits.

17.82. This situation calls for a p chart with center line $\bar{p} = \frac{6}{1000} = 0.006$ and control limits $\bar{p} \pm 3\sqrt{\frac{\bar{p}(1-\bar{p})}{350}} = 0.006 \pm 0.01238$. We take LCL $= 0$, and the UCL is 0.0184. (In order to exceed this UCL, we would need to reject at least 7 of the 350 lots.)

17.83. We find that $\bar{s} = 7.65$, so with $c_4 = 0.8862$ and $B_6 = 2.276$, we compute $\hat{\sigma} = 8.63$ and UCL $= 19.65$. One point (from sample #1) is out of control. (And, if that cause were determined and the point removed, a new chart would have s for sample #10 out of control.) The second (lower) UCL line on the control chart is the final UCL, after removing both of those samples (per the instructions in Exercise 17.84).



17.84. Without samples 1 and 10, $\bar{s} = 6.465$, $\hat{\sigma} = \bar{s}/c_4 \doteq 7.295$, and the new UCL is $2.276\hat{\sigma} = 16.60$; this line is shown on the control chart in the solution to the previous problem. Meanwhile, $\bar{\bar{x}} = 834.5$, and the control limits are $\bar{\bar{x}} \pm 3\hat{\sigma}/\sqrt{3} = 821.86$ to 847.14. The $\bar{x}$ chart gives no indication of trouble—the process seems to be in control.



17.85. **(a)** As was found in the previous exercise, $\hat{\sigma} = \bar{s}/c_4 \doteq 7.295$. Therefore, $C_p = \frac{50}{6\hat{\sigma}} \doteq 1.1423$. This is a fairly small value of C_p; the specification limits are just barely wider than the $6\hat{\sigma}$ width of the process distribution, so if the mean wanders too far from 830, the capability will drop. **(b)** If we adjust the mean to be close to 830 mm $\times 10^{-4}$ (the center of the specification limits), we will maximize C_{pk}. C_{pk} is more useful when the mean is not in the center of the specification limits. **(c)** The value of $\hat{\sigma}$ used for determining C_p was estimated from the values of s from our control samples. These are for estimating short-term variation (within those samples) rather than the overall process variation. To get a better estimate of the latter, we should instead compute the standard deviation s of the *individual* measurements used to obtain the means and standard deviations given in Table 17.11 (specifically, the 60 measurements remaining after dropping samples 1 and 10). These numbers are not available. (See "How to cheat on C_{pk}" on page 17–44 of Chapter 17.)

17.86. About 99.94%: With $\hat{\sigma} \doteq 7.295$ and mean 830, we compute $P(805 < X < 855) = P(-3.43 < Z < 3.43) = 0.9994$.

17.87. **(a)** Use a p chart, with center line $\bar{p} = \frac{15}{5000} = 0.003$ and control limits $\bar{p} \pm 3\sqrt{\frac{\bar{p}(1-\bar{p})}{100}}$, or 0 to 0.0194. **(b)** There is little useful information to be gained from keeping a p chart: If the proportion remains at 0.003, about 74% of samples will yield a proportion of 0, and about 22% of proportions will be 0.01. To call the process out of control, we would need to see two or more unsatisfactory films in a sample of 100.

17.88. Assuming $\bar{x}$ is (approximately) Normally distributed, the probability that it would fall within the 1σ level is about 0.68, so the probability that it does this 15 times is about $0.68^{15} \doteq 0.0031$.

17.89. Several interpretations of this problem are possible, but for most reasonable interpretations, the probability is about 0.3%. From the description, it seems reasonable to assume that all three points are inside the control limits; otherwise, the one-point-out rule would take effect. Furthermore, the phrase "two out of three" could be taken to mean either "*exactly* two out of three," or "*at least* two out of three." (Given what we are trying to detect, the latter makes more sense, but students may have other ideas.)

For the kth point, we name the following events:

- A_k = "that point is no more than $2\sigma/\sqrt{n}$ from the center line,"

- B_k = "that point is 2 to 3 standard errors from the center line."

For an in-control process, $P(A_k) = 95\%$ (or 95.45%) and $P(B_k) = 4.7\%$ (or 4.28%). The first given probability is based on the 68–95–99.7 rule; the second probability (in parentheses) comes from Table A or software.

Note that, for example, the probability that the first point gives no cause for concern, but the second and third are more than $2\sigma/\sqrt{n}$ from, and on the same side of, the center line, would be:

$$\tfrac{1}{2}P(A_1 \cap B_2 \cap B_3) \doteq 0.10\% \text{ (or } 0.09\%)$$

(The factor of 1/2 accounts for the second and third points being on the same side of the center line.) If the "other" point is the second or third point, this probability is the same, so if we interpret "two out of three" as meaning "*exactly* two out of three," then the total probability is three times the above number:

$$P(\text{false out-of-control signal from an in-control process}) \doteq 0.31\% \text{ (or } 0.26\%)$$

With the (more-reasonable) interpretation "*at least* two out of three":

$$\begin{aligned} P(\text{false out-of-control signal}) &= \tfrac{1}{2}P(A_1 \cap B_2 \cap B_3) + \tfrac{1}{2}P(B_1 \cap A_2 \cap B_3) + \tfrac{1}{2}P(B_1 \cap B_2) \\ &\doteq 0.32\% \text{ (or } 0.27\%) \end{aligned}$$