SOLUTIONS TO PRACTICE PROBLEMS

CHAPTER 1

1. Researchers engage in descriptive research when the goal is simply to describe phenomena; however, when research attempts to explain something by examining the relationships between phenomena, they are typically engaging in explanatory research. Evaluation research is undertaken when the goal is to determine whether a program or policy was implemented as planned and/or whether it had the intended outcomes or impacts.
2. The goal in obtaining or selecting a sample is to select it in a way that increases the chances of this sample being representative of the entire population. Using probability sampling techniques not only serves to minimize any potential bias we may have when selecting a sample, thereby making our sample more representative of the population, but also allows us to gain access to probability theory in our data analysis. This body of mathematical theory allows us to estimate more accurately the degree of error we have when generalizing results obtained from known sample statistics to unknown population parameters.
3. a. Quantitative; interval/ratio.
   b. Quantitative; interval/ratio.
   c. Quantitative; interval/ratio.
   d. Qualitative; nominal.
   e. Qualitative; nominal.
   f. Quantitative; interval/ratio.
4. The categories of a variable measured at the ordinal level of measurement can be ordered, but the distance between the categories is not quantifiable. Categories for interval-level variables have a known and equal distance between them. In addition to this, ratio-level variables have a true zero point.
5. Arrest is the independent variable, and future drunk driving behavior is the dependent variable.
6. Gender is the independent variable, and fear is the dependent variable.
7. The numerator would be the number of victimizations against people 14 to 18 years old, and the denominator would be the total population of people 14 to 18 years old.
8. Rates allow you to make comparisons across different places and time.
9.

	f	*Proportion*	%
Less than $10	16	.029	2.9
$10–$49	39	.072	7.2
$50–$99	48	.088	8.8

	f	*Proportion*	%
\$100–\$249	86	.159	15.9
\$250–\$999	102	.188	18.8
\$1,000 or more	251	.463	46.3
	n = 542		

10. The units of analysis in the Schnapp (2015) study are cities. The units of analysis in the Rydberg and Pizarro (2014) study are homicide cases.

11. The units of analysis are states. The independent variable would likely be unemployment, and the dependent variable would be crime.

12. The units of analysis would be the police departments.

2CHAPTER 2

1. The first grouped frequency distribution is not a very good one for a number of reasons. First, the interval widths are not all the same size. Second, the class intervals are not mutually exclusive. A score of 7 could go into either the first or second interval, and a score of 10 could go into either the second or third class interval. Third, the first class interval is empty; it has a frequency of zero. Fourth, there are too few class intervals; the data are "bunched up" into only three intervals, and you do not get a very good sense of the distribution of these scores. The second grouped frequency distribution avoids all of these four problems.

2. Since "number of executions" is a quantitative continuous variable (measured at the interval/ratio level), you could use a histogram or line chart to graph the frequencies. You could not use a pie chart or bar chart because such charts are for nominal-level or ordinal-level data.

3. a. "Self-reported drug use" is measured at the ordinal level because our values consist of rank-ordered categories. We do not have interval/ratio-level measurement because although we can state that someone who reported using drugs "a lot" used drugs more frequently than someone who reported "never" using drugs, we do not know exactly how much more frequently.
   b. Since there were 30 students who reported "never" using, 150 – 30, or 120 students, must have been using drugs at some level of frequency. The ratio of users to non-users, then, is 120/30, or 4 to 1.
   c. 35/10, or 3.5 to 1.
   d. The first thing we would want to do is arrange the data in some order. Since we have ordinal-level data, we can order the categories in ascending or descending order.

Value	*f*	*p*	%
Never	30	.2000	20.00
A few times	75	.5000	50.00
More than a few times	35	.2333	23.33
A lot	10	.0667	6.67

e. Since the proportion of non-users ("never") was .20, the proportion of respondents who reported using drugs must be 1 – .20, or .80. Another way to determine this is to determine the relative frequency of users (75 + 35 + 10) / 150 = 120/150 = .80.

f. .0667 of the respondents reported using drugs "a lot."

4. a. Since these data are measured at the nominal level, there is no way to "correctly" order the values in any numerical order. For our frequency distribution, then, we can employ any ordering of the values.

Value	*f*	*p*	%
Community facility	5,428	.2616	26.16
Minimum security	3,285	.1583	15.83
Medium security	1,733	.0835	8.35
Maximum security	875	.0422	4.22
Pretrial release	9,430	.4544	45.44

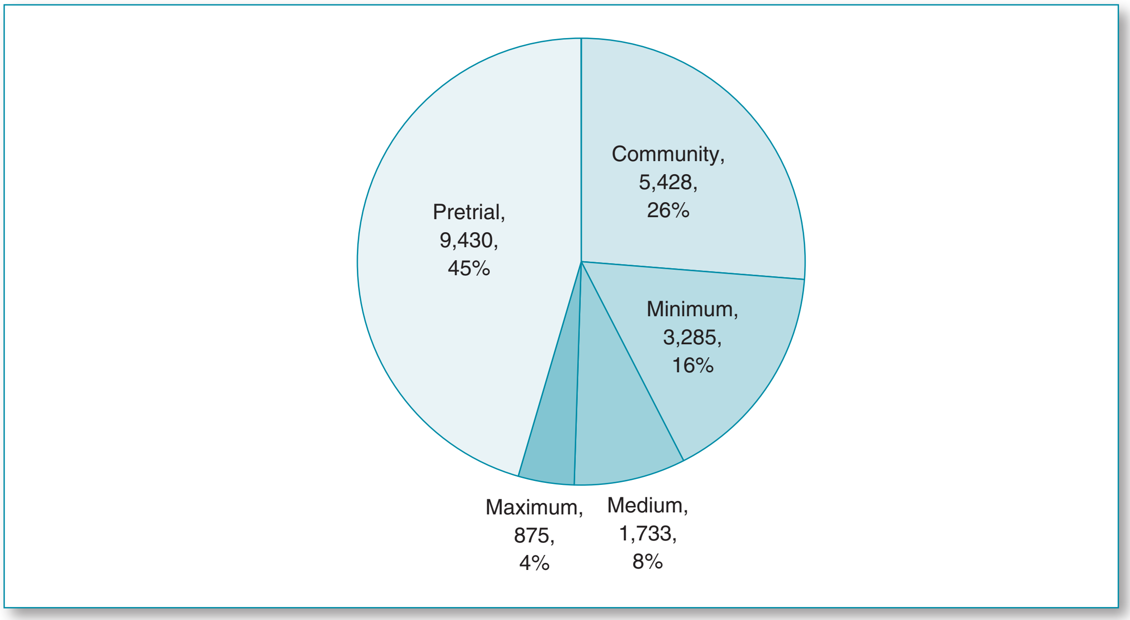
b. 3,285 / 20,751 = .1583; .1583 × 100 = 15.83%

c. 875 / 20,751 = .0422

d. 430 / 20,751 = .4544; .4544 × 100 = 45.44%

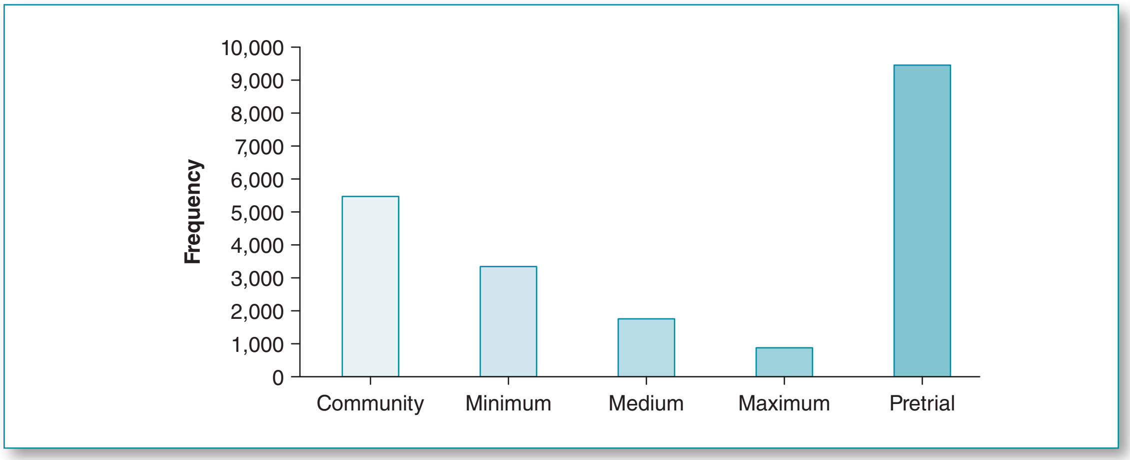
e. Since these data are measured at the nominal level, the correct graph to use would be either a pie chart or a bar chart. Several different types follow.

Pie Chart of Distribution of Correctional Facilities



Note: Percentages may not add to 100 due to rounding.

Frequency Distribution of Inmates in Correctional Institutions



5. a.

Value	f	cf	p	cp	%	c%
10	5	5	.20	.20	20	20
11	3	8	.12	.32	12	32
12	0	8	.00	.32	0	32
13	2	10	.08	.40	8	40
14	2	12	.08	.48	8	48
15	7	19	.28	.76	28	76
16	3	22	.12	.88	12	88
17	0	22	.00	.88	0	88
18	0	22	.00	.88	0	88
19	1	23	.04	.92	4	92
20	2	25	.08	1.00	8	100

b.

Value	f	p	%
Male	16	.64	64
Female	9	.36	36

c. Using the cumulative frequency column, we can determine that 10 recruits scored 13 or lower on the exam. That means that 25 – 10, or 15 recruits, must have scored 14 or higher, so 15 recruits passed the exam. Since 15 / 25 = .60, we can calculate that 60% of the recruits passed the test. We could also have used the cumulative percentage column to find this answer. Using the cumulative percentage column, we can determine that 40% of the recruits scored 13 or lower on the exam. This means that 100% – 40%, or 60%, of the recruits must have scored 14 or higher and passed the exam.

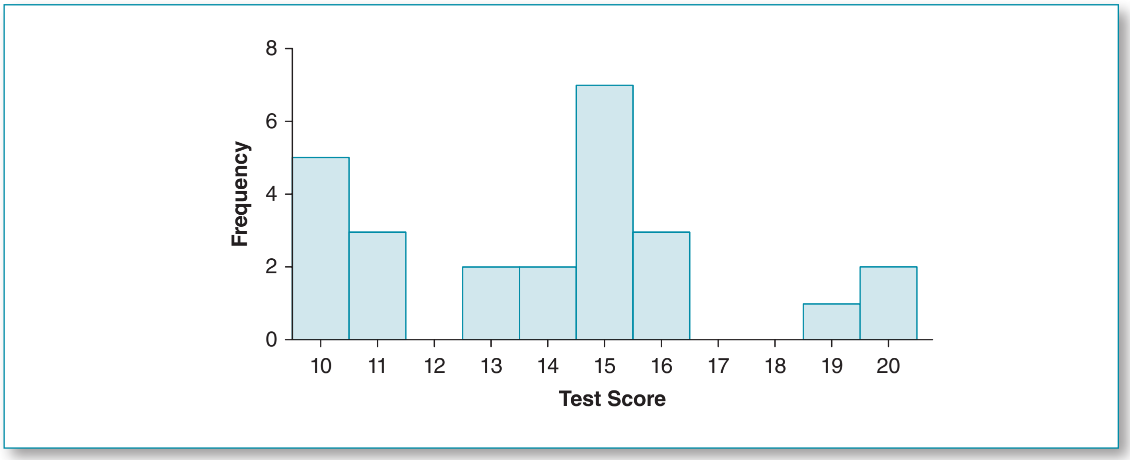
d. Of the 25 recruits, 3, or .12 (3/25), received a score of 18 or higher on the exam and "passed with honors."

e. Using the cumulative frequency column, we can easily see that 10 recruits received a score of 13 or lower on the exam.

f. In this class of recruits, 64% were male and 36% were female.

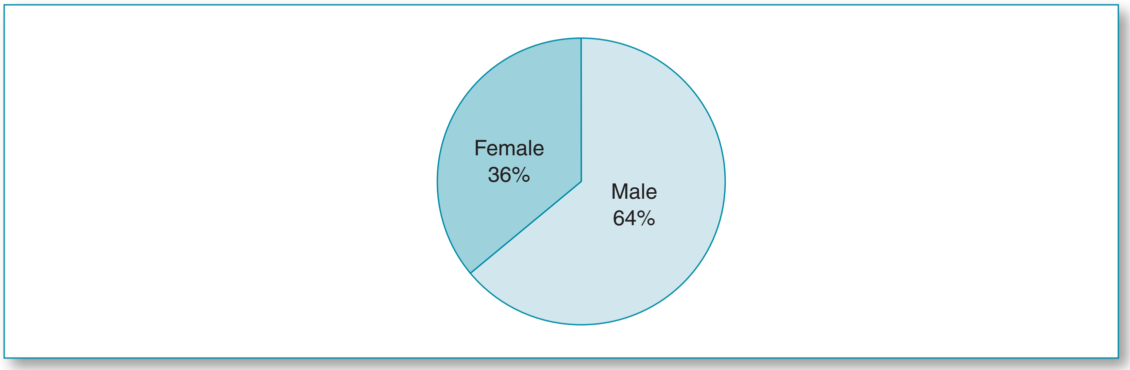
g. The test scores would have to be graphed with a histogram since the data are quantitative.

Distribution of Test Scores for Recruit Class



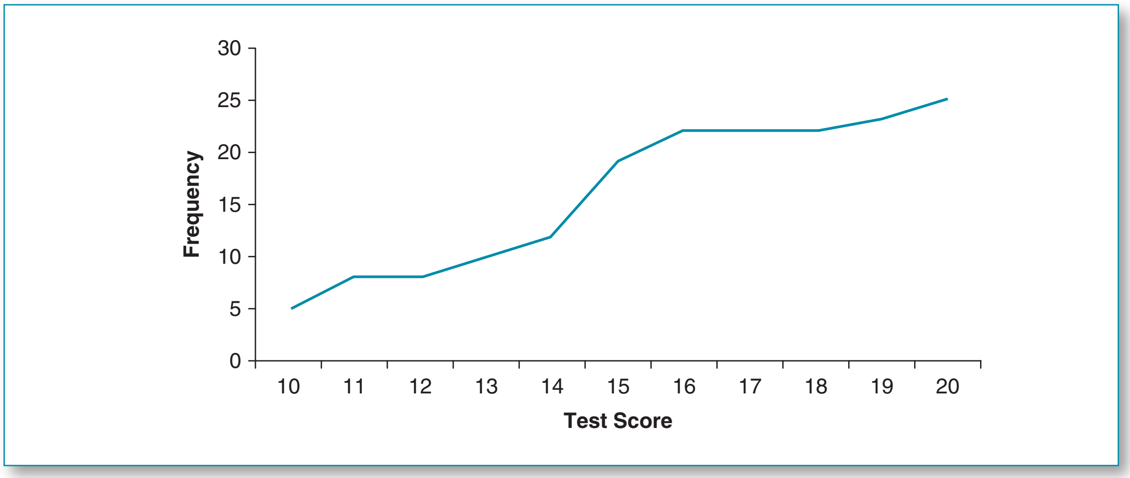
A pie chart of the percentages for the gender data would look like this:

Gender Distribution of Recruit Class



h. A cumulative frequency distribution for the test scores would look like this:

Cumulative Frequency Line Graph for Test Score Data



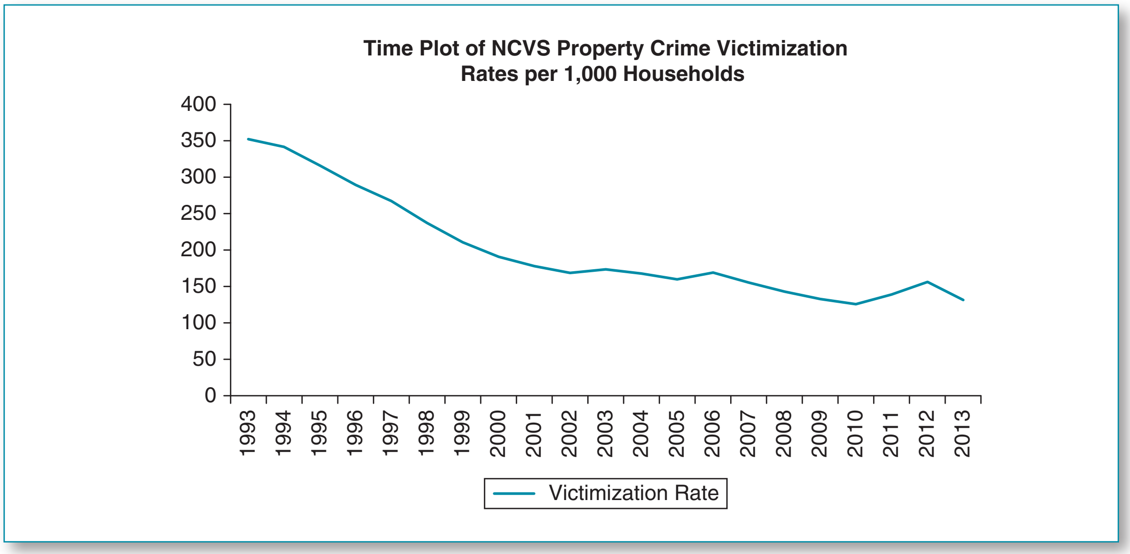
6. The grouped distribution with interval width of 6 with real limits and midpoints:
   a. The real limits are shown.
   b. The midpoints of the class intervals are shown.
   c. Looking at the cumulative frequencies, we can determine that 60 of these people reported committing their first offense before age 31. The result is equal to the sum of the frequencies for the first five class intervals.

Stated Limits	*Real Limits*	*f*	*m*	*p*	%	*cf*	*cp*	*c%*
6–10	5.5–10.5	3	8	.04	4	3	.04	4
11–15	10.5–15.5	18	13	.24	24	21	.28	28
16–20	15.5–20.5	15	18	.20	20	36	.48	48
21–25	20.5–25.5	13	23	.17	17	49	.65	65
26–30	25.5–30.5	11	28	.15	15	60	.80	80
31–35	30.5–35.5	10	33	.13	13	70	.93	93
36–40	35.5–40.5	4	38	.05	5	74	.98	98
41–45	40.5–45.5	1	43	.01	1	75	.99	99

   d. Of these persons, .24 committed their first offense between the ages of 11 and 15. It is the proportion for the second class interval.
   e. Looking at the cumulative percentage column, we can determine that 48% of the 75 persons committed their first offense at or before the age of 20, which means that 99% – 48%, or 51%, of them committed their first offense at age 20 or older. Note that the cumulative percentages do not sum exactly to 100% because of rounding error.
   f. Using the column of cumulative percentages, we can determine that 28% committed their first offense before the age of 16.

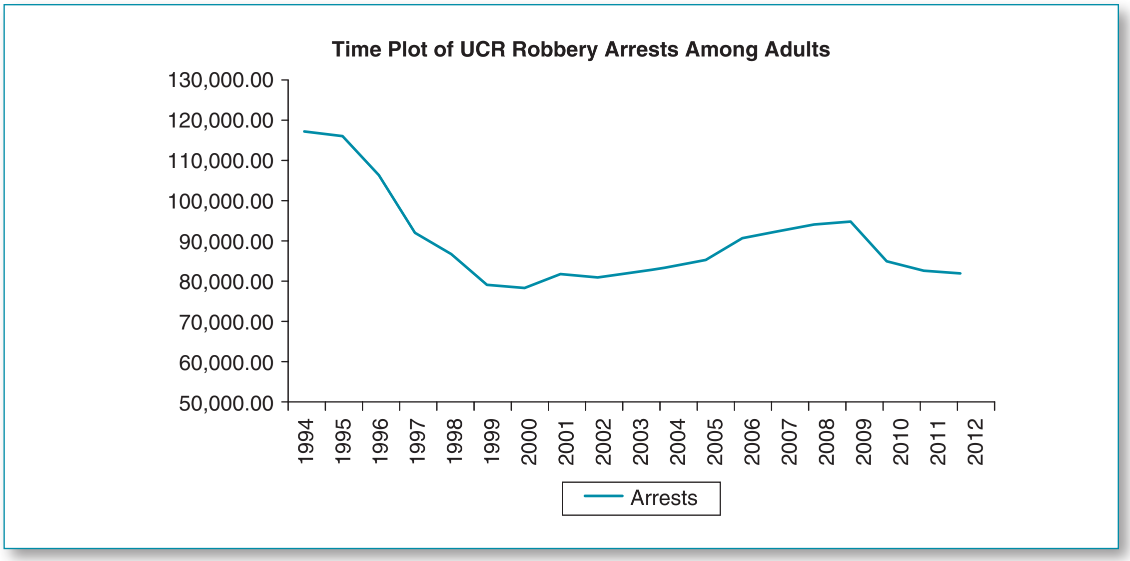
7. A time plot of the property crime victimization data from the National Crime Victimization Survey (NCVS) over the time period 1973–2013 would look like this:

Time Plot of NCVS Property Crime Victimization Rates per 1,000 Households



The time plot shows a fairly consistent downward trend in the property crime victimization rate for the duration of the period. The sharpest decline is during the 1990s. Beginning in 2001, there was a leveling off in the rate of property victimizations until 2006, after which there was another consistent decline in the property victimization rate until 2010, when there was a short increase before a final drop resulted in a property victimization rate in 2013 that was approximately 37% of what it was in 1993.

8.



CHAPTER 3

1. The mode for these data is "some friends" because this value appears more often than any other value ($f = 85$). The mode tells us that in our sample, more youths reported having "some" delinquent friends than any other possible response. We could not calculate a mean with these data because this variable is measured at the ordinal level and the mean requires data measured at the interval/ratio level. For example, although we can say that a person with "some" delinquent friends has more delinquent friends than a person with "none," we do not know exactly how many more (e.g., 1? 2? 10?). Without this knowledge, we cannot calculate the mean as a measure of central tendency.
2. To find the median salary, let's first rank-order the data from low to high:

   $25,900

   $26,100

   $27,800

   $28,400 the median salary

   $29,500

   $31,000

   $32,100

   With 7 scores, the median is the fourth score from the top or bottom, and the median salary for these correctional officers is $28,400. Note that one half of these salaries are higher than the median and one half are lower. The mean salary is

$$\bar{X} = \frac{\$25{,}900 + \$26{,}100 + \$27{,}800 + \$28{,}400 + \$29{,}500 + \$31{,}00 + \$32{,}100}{7}$$
$$\bar{X} = \$28{,}686$$

The median and mean salaries are very close to each other.

3. The best measure of central tendency for these data is probably the median. The mean would not be the best in this case because it would be inflated by the presence of a positive outlier. New Orleans, Louisiana, with a homicide rate of 43.3 per 100,000, has a homicide rate substantially higher than the other cities. When New Orleans is included in the data, the mean is equal to 10.62 homicides per 100,000, and the median is 7.25.

4. The mean is

$$\bar{X} = \frac{339}{20}$$
$$\bar{X} = 16.95$$

On average, these injection drug users committed nearly 17 crimes during the past 2 years.
The median is the average of the 10th and 11th scores in the rank-ordered frequency distribution:

$$\text{Median} = \frac{8+9}{2}$$
$$\text{Median} = 8.5$$

The median number of crimes over the past 2 years was 8.5.
The median is the preferred measure of central tendency for these data.
The value of the mean is inflated by the existence of two extremely large scores (88 and 112).

5. The most appropriate measure of central tendency for these data is the mode because the data are measured at the nominal level. The modal, or most frequent, reason for requesting the police when the subject was without mental illness was for a "potential offense."

6. The mode is the interval 30%–39% of the police officers since this interval has the highest frequency (38).

$$\text{Median} = 29.5 + \left(\frac{\frac{100+1}{2} - 44}{38} \right)(10)$$
$$\text{Median} = 31.2$$

Thus, 31.2% of the officers in the department do narcotics investigation work.

To calculate the mean with these grouped data, remember that you first must determine the midpoint of each class interval. With the midpoints, you can find the mean:

$$\bar{X} = \frac{4.5(5) + 14.5(13) + 24.5(26) + 34.5(38) + 44.5(14) + 54.5(2) + 64.5(2)}{100}$$
$$\bar{X} = 30.2$$

On average, then, approximately 30% of the officers on these city police departments do narcotics investigation work.

7. Mean number of executions:

$$\bar{X} = \frac{(42+37+52+46+43+43+39+35)}{8}$$
$$\bar{X} = 42.125 \text{ executions per year}$$

The median is the average of the fourth and fifth years in the rank-ordered frequency distribution.

$$\text{Median} = \frac{42+43}{2}$$
$$\text{Median} = 42.5 \text{ executions per year}$$

When executions for the year 2006 (53) are added to the data, the mean becomes

$$\bar{X} = \frac{390}{9} = 43.33 \text{ executions per year}$$

The median becomes 43 executions. Either the mean or the median would be appropriate here because 53 executions in 2006 is not such a large outlier.

8. Again, with grouped data, you need to find the midpoint of each class interval before you can calculate the mean. The mean is

$$\bar{X} = \frac{[.5(85)+2.5(70)+4.5(30)+6.5(15)]}{200}$$
$$\bar{X} = \frac{450}{200}$$
$$\bar{X} = 2.25 \text{ times}$$

The median is

$$\text{Median} = 2.5 + \left(\frac{\frac{201}{2} - 85}{70} \right)(2)$$
$$\text{Median} = 2.94 \text{ times}$$

The mode is equal to the interval 0–1 times because it contains the highest frequency (85).

9. The mean is equal to

$$\bar{X} = \frac{1200}{20}$$
$$\bar{X} = 60 \text{ beats per minute}$$

The median is equal to 60.5 beats per minute.

The mean and the median are very comparable to one another. This suggests that there are no or few extreme (outlying) scores in the data and that the data are not skewed.

CHAPTER 4

1. Measures of central tendency capture the most "typical" score in a distribution of scores (the most common, the score in the middle of the ranked distribution, or the average), whereas measures of dispersion capture the variability in our scores, or how they are different from each other or different from the central tendency. It is important to report both central tendency and dispersion measures for our variables because two groups of scores may be very similar in terms of their central tendency but very different in terms of how dispersed the scores are.

2. The variation ratio for those whose current offense is a property crime is

$$VR = 1 - \frac{75}{125}$$
$$VR = .40$$

So, .40 of the current offenses did not fall into the modal category.

The variation ratio for those whose current offense is a violent crime is

$$VR = 1 - \frac{50}{110}$$
$$VR = .54$$

So, .54 of the current offenses did not fall into the modal category.

The variation ratio for those whose current offense is a drug crime is

$$VR = 1 - \frac{110}{230}$$
$$VR = .52$$

So, .52 of the current offenses did not fall into the modal category.

Finally, the variation ratio for those whose offense is current offense is a status is

$$VR = 1 - \frac{320}{575}$$
$$VR = .44$$

So, .44 of the current offenses did not fall into the modal category.

The most dispersion in these nominal-level variables is for violent crimes and drug crimes. For violent crimes .54 of the cases are in a value other than the modal value, and for drug crimes .52 of the cases are in a value other than the modal value. The least dispersion in these variables occurs for property crime with a variation ratio of .40.

3. The first thing we need to do is calculate the mean. This problem will give you experience in calculating a mean for grouped data. You should find the mean equal to 8.6 prior thefts. We are now ready to do the calculations necessary to find the variance and standard deviation.

m_i	$m_i - \bar{X}$	$(m_i - \bar{X})^2$	f	$f(m_i - \bar{X})^2$
2	2 – 8.6 = –6.6	43.56	76	3,310.56
7	7 – 8.6 = –1.6	2.56	52	133.12
12	12 – 8.6 = – 3.4	11.56	38	439.28
17	17 – 8.6 = 8.4	70.56	21	1,481.76
22	22 – 8.6 = 13.4	179.56	10	1,795.60
27	27 – 8.6 = 18.4	338.56	8	2,708.48
				Σ = 9,868.80

$$s^2 = \frac{9{,}868.80}{204}$$
$$s^2 = 48.38$$

The variance is equal to 48.38.

$$s = \sqrt{\frac{9{,}868.80}{204}}$$
$$s = 6.95$$

The standard deviation is equal to 6.95.

4. To answer these questions, let's first rank-order the data:

5 9
5 9
6 10
7 10
8 10
9 10
9 11
9 11
9 12
9 12

a. The range is 12 – 5 = 7 years of education.
b. The median position is (20 + 1) / 2 = 10.5, so the truncated median position is equal to 10. The position of the quartiles, then, is (10 + 1) / 2 = 5.5. With this, we can identify the first quartile as equal to 8.5 and the third quartile as equal to 10, so the interquartile range is IQR = 10 – 8.5 = 1.5.

c. To calculate the variance and standard deviation, we first have to determine the mean. The mean is equal to 9 years of education. Here are the calculations necessary to determine the variance and standard deviation:

x_i	$x_i - \bar{X}$	$(x_i - \bar{X})^2$
11	11 – 9 = 2	4
8	8 – 9 = –1	1
12	12 – 9 = 3	9
9	9 – 9 = 0	0
9	9 – 9 = 0	0
9	9 – 9 = 0	0
10	10 – 9 = 1	1
10	10 – 9 = 1	1
10	10 – 9 = 1	1
11	11 – 9 = 2	4
9	9 – 9 = 0	0
9	9 – 9 = 0	0
9	9 – 9 = 0	0
9	9 – 9 = 0	0
5	5 – 9 = –4	16
9	9 – 9 = 0	0
7	7 – 9 = –2	4
6	6 – 9 = –3	9
10	10 – 9 = 1	1
12	12 – 9 = 3	9
9	9 – 9 = 0	0
5	5 – 9 = –4	16
		Σ = 76

$$s^2 = \frac{76}{19}$$

$$s^2 = 4$$

The variance is equal to 4.

$$s = \sqrt{\frac{76}{19}}$$

$$s = 2$$

The standard deviation is equal to 2.

5. Let's calculate the variation ratio for each of the 3 years:

$$VR_{1980} = 1 - \frac{852}{1{,}723}$$
$$VR_{1980} = .50$$
$$VR_{1990} = 1 - \frac{979}{2{,}161}$$
$$VR_{1990} = .55$$
$$VR_{2000} = 1 - \frac{1{,}211}{3{,}202}$$
$$VR_{2000} = .62$$
$$VR_{2010} = 1 - \frac{1{,}300}{3{,}612}$$
$$VR_{2010} = .64$$

The variation ratio is consistently increasing from 1980 to 2010, which tells us that the dispersion in the nominal-level data is increasing. Practically, this says that the racial heterogeneity of the penitentiary is increasing over time.

6. Let's rank-order these data first:

5
6
7
13
17
17
22
25
29
29
30
32
32
33
41
51
56
70

a. Since the highest robbery arrest rate is for New York at 70 and the lowest is for Wyoming at 5, the range is 70 – 5 = 65.

b. The position of the median is (18 + 1) / 2 = 9.5, so the truncated median position is 9.0. We can now determine the first and third quartile positions to be (9 + 1) / 2 = 5. The third quartile score is 33, and the first quartile score is 17. The interquartile range, therefore, is 33 – 17 = 16.

c. Here are the calculations necessary to determine the variance and standard deviation. First, practice your mean calculation skills. The mean robbery arrest rate per 100,000 for this sample of 18 cities is 29. (Actually, it's 28.61, but let's make our calculations easy and round it to 29 per 100,000.)

x	x – 29	(x – 29)²
29	0	0
22	–7	49
17	–12	144
32	3	9
6	–23	529
29	0	0
17	–12	144
56	27	729
33	4	16
70	41	1681
41	12	144
7	–22	484
30	1	1
51	22	484
32	3	9
25	–4	16
13	–16	256
5	–24	576
		Σ = 5,271

So, the variance would be

$$s^2 = \frac{5{,}271}{17}$$

$$s^2 = 310.06$$

The standard deviation would be

$$s = \sqrt{\frac{5{,}271}{17}}$$

$$s = 17.61$$

CHAPTER 5

1. a. P(x = \$30,000) = 16/110 = .145

   b. P(x = \$35,000) = 7/110 = .064

   c. Yes, they are mutually exclusive events because a person cannot simultaneously have a starting salary of both \$30,000 and \$35,000. There is no joint probability of these two events.

   d. P(x ≥ \$31,000) = (19/110) + (12/110) + (15/110) + (8/110) + (7/110) = (61/110) = .555

   e. There are two ways to calculate this probability. First: P(x ≤ \$30,000) = (16/110) + (10/110) + (9/110) + (8/110) + (6/110) = (49/110) = .445. Or you can recognize that this event is the complement of the event in part d above and calculate the probability as 1 – .555 = .445.

   f. P(x = \$28,000 or \$30,000 or \$31,500 or \$32,000 or \$32,500) = (10/110) + (16/110) + (19/110) + (12/110) + (15/110) = (72/110) = .655

   g. P(x < \$25,000) = 0

   h. P(x = \$28,000 or \$32,000 or \$35,000) = (10/110) + (12/110) + (7/110) = (29/110) = .264

2. The probability and cumulative probability of 0 to 10 acquittals if the probability of an acquittal is .40 are shown below.

# of Acquittals	*p*	*cp*
0	.0060	1.0000
1	.0403	.9940
2	.1209	.9537
3	.2150	.8328
4	.2508	.6178
5	.2007	.3670
6	.1115	.1663
7	.0425	.0548
8	.0106	.0123
9	.0016	.0017
10	.0001	.0001

The probability of getting 7 or more acquittals if the true probability of an acquittal is .40, then, is .0548. To test our hypothesis, we will do the five steps:

Step 1: H_0: P(acquittal with a public defender is equal to .40)

H_1: P(acquittal with a public defender is > .40)

Step 2: The test statistic is a binomial statistic with a binomial probability distribution.

Step 3: Our alpha level is .05. We will reject the null hypothesis if the probability of observing 7 or more acquittals out of 10 cases is .05 or lower.

Step 4: We calculated the full probability distribution above. The probability of observing 7 or more acquittals out of 10 cases when the probability of an acquittal is .40 is equal to .0548.

Step 5: Since our observed probability is greater than .05, we will fail to reject the null hypothesis. Defendants with public defenders are no more likely to be acquitted than are defendants with other types of lawyers.

3. a. $z = 1.5$
   b. $z = -1.7$
   c. $z = -3.0$
   d. .0548, or slightly more than 5%, of the cases have an IQ score above 115.
   e. .6832
   f. A raw score of 70 corresponds to a z score of −3.0. The probability of a z score less than or equal to −3.0 is .001.
   g. A raw score of 125 corresponds to a z score of 2.5. The probability of a z score greater than or equal to 2.5 is .006.
4. a. P(deterred) = 80 / 120 = .67
   b. This is an unconditional probability. It is the probability that someone was deterred in the entire sample. It is not based on a prior condition.
   c. P(not deterred) = 1 − .67 = .33 since this is the complement of the event in part a above, or 40 / 120 = .33.
   d. P(impulsive) = 30 / 120 = .25
   e. These are not mutually exclusive events because it is possible to be both impulsive and not deterred. P(impulsive or not deterred) = (30 / 120) + (40 / 120) − (25 / 120) = (45 / 120) = .38.
   f. P(not deterred | impulsive) = (25 / 30) = .83
   g. P(deterred | not impulsive) = (75 / 90) = .83
   h. To determine whether being deterred and impulsivity are independent events, let's compare the unconditional probability of being deterred against the conditional probability of being deterred given that a person was not impulsive and then given that a person was impulsive.

      P(deterred) = .67

      P(deterred | not impulsive) = (75 / 90) = .83

      P(deterred | impulsive) = (5 / 30) = .17

      The unconditional probability is not comparable to the conditional probabilities; it looks like the probability that someone would be deterred by punishment depends on whether or not he or she is impulsive. The person is far more likely to be deterred by punishment if he or she is not impulsive. These are not independent events.
   i. P(impulsive and not deterred) = P(impulsive) × P(not deterred | impulsive) = (30 / 120) × (25 / 30) = (.25) × (.83) = .21
   j. P(not impulsive and deterred) = P(not impulsive) × P(deterred | not impulsive) = (90 / 120) × (75 / 90) = (.75) × (.83) = .62
5. a. A raw score of 95 is better than .9332, or 93%, of the scores. It is not in the top 5%, however, so this candidate would not be accepted.
   b. A raw score of 110 is better than .9986, or 99%, of the scores. It is in the top 5%, and this candidate would be accepted.
   c. A z score of 1.65 or higher is better than 95% of the scores. The z score of 1.65 corresponds to a raw score of 96.5, and that is the minimum score you need to get accepted.
6. A population is the universe of cases about which we would like to have information and make inferences. A population is generally quite large, consisting of many elements. An example would be the population of all adolescents in the United States. The population has characteristics (such as the mean or proportion of some variable) that are not usually known but are knowable. A sample is a subset of a population and consists of many fewer elements.

We usually take a sample from a population and use the information from the sample, which we know, to make an inference about some unknown population characteristic or parameter. An example would be a sample of 100 adolescents that we might take from the population of all adolescents in the United States. The population would have a distribution of some characteristic—let's say the number of delinquent acts committed. This characteristic would have a mean (μ) and a standard deviation (σ). The sample we drew of 100 adolescents would also have a distribution with a sample mean $(\bar{x})$ and a sample standard deviation (s). A sampling distribution is a theoretical probability distribution. It is a distribution of an infinite number of samples. For example, if we were to take an infinite number of samples of size 100 from our population of adolescents, and if for each sample of size 100 we calculated a mean, we would have a distribution of an infinite number of sample means. This theoretical probability distribution of sample means would also have a mean (μ) and a standard deviation ($\sigma / \sqrt{n}$).

7. a. The area to the right of a z score of 1.65 is equal to .0495.
   b. The area to the left of a z score of –1.65 is equal to .0495.
   c. The area either to the left of a z score of –1.65 or to the right of a z score of 1.65 is equal to .099.
   d. The area to the right of a z score of 2.33 is .0099.

8. a. P(no violence) = (60 / 250) = .24
   b. P(guards only) = (55 / 250) = .22
   c. Since these are mutually exclusive events, P(metal detectors or guards only) = (60 / 250) + (55 / 250) = .46.
   d. P(no measures) = (80 / 250) = .32
   e. Since P(guards and metal detectors or 1–4 violent acts) are not mutually exclusive acts, (55 / 250) + (75 / 250) – (15 / 250) = (115 / 250) = .46.
   f. Since P(used metal detectors only or 5+ violent acts) are not mutually exclusive acts, (60 / 250) + (115 / 250) – (30 / 250) = (145 / 250) = .58.
   g. P(no violence | no preventive measures) = 5 / 80 = .06
   h. P(no violence | some prevention)

$$\frac{10+15+30}{60+55+55} = \frac{55}{170} = .32$$

   i. P(5+ violent acts | metal detectors only) = 30 / 60 = .50
   j. P(5+ violent acts | guards and metal detectors) = 10 / 55 = .18
   k. Let's calculate the unconditional probability that the school would have 5 or more violent acts:

$$P(5+\text{ violent acts}) = 115/250 = .46$$

Now let's calculate the conditional probability of 5 or more violent acts at each different level of the independent variable:

P(5+ violent acts | no preventive measures) = 50/80 = .62, P(5+ violent acts | metal detectors only) = 30/60 = .50, P(5+ violent acts | guards only) = 25/55 = .45, P(5 + violent acts | guards and metal detectors) = 10/55 = .18

The conditional probabilities are not comparable to the unconditional probability of 5 or more violent acts. In fact, it looks as if the type of preventive measure a school uses is related to the probability that 5 or more violent acts were committed. These are not independent events.

   l. P(no violent acts and guards only) = P(no violent acts) × P(guards only | no violent acts) = (60/250) × (15/60) = .06

m. P(no preventive measures and 5+ violent acts) = P(no preventive measures) × P(5+ violent acts | no preventive measures) = (80 / 250) × (50 / 80) = .20

n. P(guards together with metal detectors and 1–4 violent acts) = P(guards and metal detectors) × P(1–4 violent acts | guards and metal detectors) = (55 / 250) × (15 / 55) = .06

9. a. To see how unusual 9 prior arrests are in this population, let's transform the raw score into a z score:

$$z = \frac{9-6}{2} = 1.50.$$

Taking a z score of 1.50 to the z table, we can see that the area to the right of this score comprises approximately 7% of the area of the normal curve. Those who have 9 prior arrests, then, are in the top 7% of this population. Since they are not in the top 5%, we would not consider them unusual.

b. A raw score of 11 prior arrests corresponds to a z score of

$$z = \frac{11-6}{2} = 2.50.$$

A z score of 2.50 is way at the right or upper end of the distribution. z scores of 2.50 or greater are greater than approximately 99% of all the other scores. This person does have an unusually large number of prior arrests since they are in the top 5%.

c. A raw score of 2 prior arrests corresponds to a z score of

$$z = \frac{2-6}{2} = -2.0.$$

A z score of –2.0 falls lower than almost 98% of all the other scores. The person with only 2 prior arrests, then, does have an unusually low number for this population since he or she is in the bottom 5%.

10. The central limit theorem is a statistical proposition that holds that if an infinite number of random samples of size n are drawn from *any* population with mean μ and standard deviation σ, then as the sample size becomes large, the sampling distribution of sample means will become normal with mean μ and standard deviation σ/n. The central limit theorem enables us to make three assumptions about sampling distributions when the sample size is large: (a) we can assume that the mean of the sampling distribution is equal to the population mean, μ; (b) we can assume that the standard deviation of the sampling distribution is equal to $\sigma/\sqrt{n}$; (c) we can assume that the sampling distribution is normally distributed even if the population from which the sample was drawn is not.

CHAPTER 6

1. The purpose of confidence intervals is to give us a range of values for our estimated population parameter rather than a single value or a point estimate. The estimated confidence interval gives us a range of values within which we believe, with varying degrees of confidence, that the true population value falls. The advantage of providing a range of values for our estimate is that we will be more likely to include the population parameter. Think of trying to estimate your final exam score in this class. You are more likely to be accurate if you are able to estimate an interval within which your actual score will fall, such as "somewhere between 85 and 95," than if you have to give a single value as your estimate, such as "it will be an 89." Note that the wider you make your interval (consider "somewhere between 40 and 95"), the more accurate you are likely to be in that your exam score will probably fall within that very large interval. However, the price of this accuracy is precision; you are not being very precise in estimating that your final exam score will be between 40 and 95. In this case, you will be very confident but not

very precise. Note also that the more narrow or precise your interval is, the less confident you may be about it. If you predicted that your final exam score would be between 90 and 95, you would be very precise. You would probably also be far less confident of this prediction than of the one where you stated your score would fall between 40 and 95. Other things (such as sample size) being equal, there is a trade-off between precision and confidence.

2. At small sample sizes, the *t* distribution is flatter than a *z* distribution and has flatter tails on both ends of the distribution. When the sample size is 120 or more, the two distributions are virtually identical. If the population does not depart dramatically from normality, we can use the *z* distribution with sample sizes of 30 or more.

3.

$$95\% \text{ c.i.} = 4.5 \pm 1.96\left(\frac{3.2}{\sqrt{110-1}}\right)$$
$$95\% \text{ c.i.} = 4.5 \pm 1.96\left(\frac{3.2}{10.44}\right)$$
$$95\% \text{ c.i.} = 4.5 \pm 1.96(.31)$$
$$95\% \text{ c.i.} = 4.5 \pm .61$$
$$3.89 \leq \mu \leq 5.11$$

We are 95% confident that the mean level of marijuana use in our population of teenagers is between 3.89 and 5.11 times per year. This means that if we were to take an infinite number of samples of size 110 from this population and estimate a confidence interval around the mean for each sample, 95% of those confidence intervals would contain the true population mean.

4.

$$95\% \text{ c.i.} = 4.5 \pm 2.0\left(\frac{3.2}{\sqrt{25-1}}\right)$$
$$95\% \text{ c.i.} = 4.5 \pm 2.0\left(\frac{3.2}{4.90}\right)$$
$$95\% \text{ c.i.} = 4.5 \pm 2.0(.65)$$
$$95\% \text{ c.i.} = 4.5 \pm 1.30$$
$$3.20 \leq \mu \leq 5.80$$

Our confidence interval is much wider when our sample size is 25 than when it is 110. This is because with a smaller sample size, our sampling error becomes greater and we have to use the *t* probability distribution rather than the *z* probability distribution. Because the standard deviation of the sampling distribution is a function of sample size, it increases whenever the sample size (n) decreases. When our sample size was reduced from 110 to 25, the standard deviation of the sampling distribution increased from .31 to .65. The increase in the standard deviation of the sampling distribution (standard error) increased the width of our interval. With a sample size of 25, we are 95% confident that the true population mean is between 3.20 and 5.80 times per year.

5. The standard deviation of the sampling distribution is the standard deviation of an infinite number of sample estimates [means ($\overline{X}$) or proportions (p)], each drawn from a sample with sample size equal to n. It is also called the standard error. The sample size affects the value of the standard error (see problems 3 and 4 above). At a fixed confidence level, increasing the sample size will reduce the size of the standard error and, consequently, the width of the confidence interval.

6. Because we have a small sample ($n = 20$), we have to use the t distribution to build our 99% confidence interval around the sample mean. Therefore, we go to the t table (Table B.3) to find our t value, with $n - 1$, or 19, degrees of freedom. We hope that, in finding the correct t value from the table, you remembered that confidence interval problems are always two-tailed problems since you cannot be certain whether your point estimate overestimates or underestimates the true population value. The critical value of t with 19 degrees of freedom and alpha = .01 is equal to 2.861. The 99% confidence interval would be

$$\begin{aligned}
99\%\ \text{c.i.} &= 18 \pm 2.861\left(\frac{4}{\sqrt{20-1}}\right)\\
99\%\ \text{c.i.} &= 18 \pm 2.861\left(\frac{4}{4.36}\right)\\
99\%\ \text{c.i.} &= 18 \pm 2.861(.92)\\
99\%\ \text{c.i.} &= 18 \pm 2.63\\
15.37 &\le \mu \le 20.63
\end{aligned}$$

7. To find a 95% confidence interval around a sample mean of 560 with a standard deviation of 45 and a sample size of 15, you would have to go to the t table. With $n = 15$, there are 14 degrees of freedom. Since confidence intervals are two-tailed problems, the value of t you should obtain is 2.145. Now you can construct the confidence interval:

$$\begin{aligned}
95\%\ \text{c.i.} &= 560 \pm 2.145\left(\frac{4.5}{\sqrt{15-1}}\right)\\
95\%\ \text{c.i.} &= 560 \pm 2.145\left(\frac{45}{3.74}\right)\\
95\%\ \text{c.i.} &= 560 \pm 2.145(12.03)\\
95\%\ \text{c.i.} &= 560 \pm 25.8\\
534.2 &\le \mu \le 585.8
\end{aligned}$$

You can say that you are 95% confident that the true police response time is between 534 seconds (almost 9 minutes) and 586 seconds (almost 10 minutes).

8. Since we can assume that the sample size is large enough ($286 \times .44 > 5$ and $286 \times .56 > 5$), the 95% confidence interval for the traditional treatment group is found with the z distribution:

$$\begin{aligned}
95\%\ \text{c.i.} &= .44 \pm 1.96\sqrt{\frac{.44(1-.44)}{286}}\\
95\%\ \text{c.i.} &= .44 \pm 1.96\sqrt{\frac{.25}{286}}\\
95\%\ \text{c.i.} &= .44 \pm 1.96\sqrt{.001}\\
95\%\ \text{c.i.} &= .44 \pm 1.96(.032)\\
95\%\ \text{c.i.} &= .44 \pm .06\\
95\%\ \text{c.i.} &= .38\ \text{to}\ .50, \text{or}\ 38\%\ \text{to}\ 50\%\\
.38 &\le P \le .50
\end{aligned}$$

We can also assume a large sample for the Paint Creek Youth Group (318 × .51 > 5 and 318 × .49 > 5), and the 95% confidence interval is

$$95\%\ \text{c.i.} = .38 \pm 1.96\sqrt{\frac{.38(1-.38)}{318}}$$
$$95\%\ \text{c.i.} = .38 \pm 1.96\sqrt{\frac{.24}{318}}$$
$$95\%\ \text{c.i.} = .38 \pm 1.96\sqrt{.001}$$
$$95\%\ \text{c.i.} = .38 \pm 1.96(.032)$$
$$95\%\ \text{c.i.} = .38 \pm .06$$
$$95\%\ \text{c.i.} = .32 \text{ to } .44, \text{ or } 32\% \text{ to } 44\%$$
$$.32 \leq P \leq .44$$

9. When we increased the confidence interval from 95% to 99%, we would see that the width of the confidence interval would also increase. This is because being more confident that our estimated interval contains the true population parameter (99% confident as opposed to 95% confident) comes at the price of a wider interval (all other things being equal). You should remember, from the discussion in this chapter, that you can increase the level of your confidence without expanding the width of the interval by increasing your sample size.

2CHAPTER 7

1. The z test and z distribution may be used for making one-sample hypothesis tests involving a population mean under two conditions: (1) if the population standard deviation (σ) is known and (2) if the sample size is large enough ($n > 30$) so that the sample standard deviation (s) can be used as an unbiased estimate of the population standard deviation. If either of these two conditions is not met, hypothesis tests about one population mean must be conducted with the t test and t distribution.
2. In our first hypothesis test, the null and alternative hypotheses would be

$$H_0: \mu = \$2{,}222$$
$$H_1: \mu \neq \$2{,}222$$

If we believed the dollar amount lost by burglary victims to be *higher* than \$2,222, our null hypothesis would be the same, but we would assume the following about the alternative hypothesis: $H_1: \mu > \$2{,}222$.

3. The null and alternative hypotheses are

$$H_0: \mu = 4.6$$
$$H_1: \mu < 4.6$$

Since our sample size is large ($n > 30$), we should use the z test and z distribution. With an alpha of .01 and a one-tailed test, our critical value of z is 2.33. Our decision rule is to reject the null hypothesis if our obtained value of z is 2.33 or greater (reject H_0 if $z_{obt} > 2.33$). The value of z_{obt} is

$$z_{obt} = \frac{6.3 - 4.6}{1.9 / \sqrt{64}}$$
$$z_{obt} = \frac{1.7}{.2375}$$
$$z_{obt} = 7.16$$

Because 7.16 is greater than the critical value of 2.33 and falls in the critical region, we will reject the null hypothesis that the population mean is equal to 4.6 times.

4. The null and alternative hypotheses are

$$H_0: \mu = 3.5$$

$$H_1: \mu < 3.5$$

With an alpha of .05, a one-tailed test, and 28 degrees of freedom, our critical value of t is –1.701. Our decision rule is to reject the null hypothesis if our obtained value of t is –1.701 or less. The value of t_{obt} is

$$t_{obt} = \frac{2.9 - 3.5}{.7 / \sqrt{31}}$$
$$t_{obt} = \frac{-.6}{.7 / 5.57}$$
$$t_{obt} = \frac{-.6}{.126}$$
$$t_{obt} = -4.76$$

Because this is lower than the critical value of –1.701 and falls in the critical region, we will reject the null hypothesis that the population mean is equal to 3.5 acts of vandalism.

5. The null and alternative hypotheses are

$$H_0: \mu = 25.9$$

$$H_1: \mu \neq 25.9$$

The critical value of z is ±2.58. The value of z_{obt} is

$$z_{obt} = \frac{27.3 - 25.9}{6.5/\sqrt{175}}$$
$$z_{obt} = \frac{1.40}{6.5/13.23}$$
$$z_{obt} = \frac{1.40}{.49}$$
$$z_{obt} = 2.86$$

Because this value is greater than the critical value of 2.58 and falls in the critical region, we will reject the null hypothesis that the population mean is equal to 25.9 months.

6. The null and alternative hypotheses are

$$H_0: \mu = 15$$

$$H_1: \mu \neq 15$$

Because our sample size is small, we must use the t test and the t distribution. With 14 degrees of freedom, an alpha of .05, and a two-tailed test, our critical value of t is ±2.145. The value of t_{obt} is

$$t_{obt} = \frac{16.4 - 15}{4/\sqrt{15}}$$
$$t_{obt} = \frac{1.40}{4/3.87}$$
$$t_{obt} = \frac{1.40}{1.03}$$
$$t_{obt} = 1.36$$

Because this value is not greater than the critical value of 2.145 and does not fall in the critical region, we fail to reject the null hypothesis that the population mean is equal to 15 hours.

7. The null and alternative hypotheses are

$$H_0: \mu = 4$$

$$H_1: \mu > 4$$

With 11 degrees of freedom, an alpha of .01, and a one-tailed test, our critical value of t is 2.718. Our decision rule is to reject the null hypothesis if our obtained value of $t > 2.718$. The value of t_{obt} is

$$t_{obt} = \frac{6.3 - 4}{1.5/\sqrt{12}}$$
$$t_{obt} = \frac{2.30}{1.5/3.46}$$
$$t_{obt} = \frac{2.30}{.43}$$
$$t_{obt} = 5.35$$

Because this value is greater than the critical value of 2.718 and falls in the critical region, we decide to reject the null hypothesis that the population mean is equal to 4 arrests.

8. The null and alternative hypotheses are

$$H_0: \mu = 25$$

$$H_1: \mu < 25$$

Our decision rule is to reject the null hypothesis if our obtained value of $t < -1.729$. The value of t_{obt} is

$$t_{\text{obt}} = \frac{23-25}{6/\sqrt{20}}$$

$$t_{\text{obt}} = \frac{-2}{6/4.47}$$

$$t_{\text{obt}} = \frac{-2}{1.34}$$

$$t_{\text{obt}} = -1.49$$

Because this value is not less than the critical value of –1.729 and does not fall in the critical region, we fail to reject the null hypothesis that the population mean is equal to 25 minutes.

9. The null and alternative hypotheses are

$$H_0: P = .45$$

$$H_1: P < .45$$

Because this is a problem involving a population proportion with a large sample size *(n* = 200), we can use the z test and the z distribution. Our decision rule is to reject the null hypothesis if our obtained value of z is –2.33 or less. The value of z_{obt} is

$$z_{\text{obt}} = \frac{.23-.45}{\sqrt{\frac{.45(.55)}{200}}} = -6.25$$

Because our obtained value of z is less than the critical value of -2.33 and falls in the critical region, we will reject the null hypothesis that the population proportion is equal to .45, or 45%.

10. The null and alternative hypotheses are

$$H_0: P = .20$$

$$H_1: P \neq .20$$

Because this is a problem involving a population proportion with a large sample size ($n = 60; 60 \times .2 > 5$), we can use the z test and the z distribution. Our decision rule is to reject the null hypothesis if our obtained value of $z_{\text{obt}} \leq -1.96$ or $z_{\text{obt}} \geq 1.96$. The value of z_{obt} is

$$z_{\text{obt}} = \frac{.31-.20}{\sqrt{\frac{.20(.80)}{60}}} = 2.13$$

Because this is greater than the critical value of 1.96 and falls in the critical region, we decide to reject the null hypothesis that the population proportion is equal to .20, or 20%, of the homes.

11. The null and alternative hypotheses are

$$H_0: P = .31$$

$$H_1: P > .31$$

Because this is a problem involving a population proportion with a large sample size ($n = 110$), we can use the z test and the z distribution. Our decision rule is to reject the null hypothesis if our obtained value of z is 1.65 or greater. The value of z_{obt} is

$$z_{\text{obt}} = \frac{.46 - .31}{\sqrt{\dfrac{.31(.69)}{110}}} = -3.40$$

Because this is greater than the critical value of 1.65 and falls in the critical region, we will reject the null hypothesis that the population proportion is equal to .31, or 31%.

CHAPTER 8

1. a. The type of institution is the independent variable, and satisfaction with one's job is the dependent variable.
   b. There are a total of 185 observations.
   c. There are 115 persons who reported that they were not satisfied with their jobs and 70 persons who were satisfied with their jobs.
   d. There were 45 people working in medium-security institutions and 140 people employed in maximum-security institutions.
   e. This is a 2 × 2 contingency table.
   f. A total of 30 correctional officers are in medium-security institutions and like their jobs.
   g. A total of 100 correctional officers are in maximum-security institutions and do not like their jobs.
   h. There is (2 – 1) (2 – 1) or 1 degree of freedom.
   i. The risk of not being satisfied with your job is

$$.33\left(\frac{15}{45}\right)$$

   in medium-security institutions and

$$.71\left(\frac{100}{140}\right)$$

   in maximum-security institutions.

   It looks as if the type of institution one works in is related to one's job satisfaction. Officers are far more likely to be dissatisfied with their jobs if they work in a maximum-security facility than if they work in a medium-security facility.

   j. **Step 1:** H_0: Type of institution and level of job satisfaction are independent.

   H_1: Type of institution and level of job satisfaction are not independent.

   Step 2: Our test statistic is a chi-square test of independence, which has a chi-square distribution.

   Step 3: With 1 degree of freedom and an alpha of .05, our $\chi^2_{\text{crit}} = 16.812$. The critical region is any obtained chi-square to the right of this. Our decision rule is to reject the null hypothesis when $\chi^2_{\text{obt}} \geq 3.841$.

Step 4: When we calculate our obtained chi-square, we find that it is $\chi^2_{obt} \geq 16.812$.

Step 5: With a critical value of 3.841 and an obtained chi-square statistic of 21.11, our decision is to reject the null hypothesis. Our conclusion is that type of institution employed at and job satisfaction for a correctional officer are not independent; there is a relationship between these two variables in the population.

We could use several different measures of association for a 2 × 2 contingency table. Our estimated value of Yule's *Q* would be –.67, which would tell us that there is a strong negative relationship between type of institution and job satisfaction. More specifically, we would conclude that those who work in a maximum-security facility have less job satisfaction. Since we have a 2 × 2 table, we could also have used the phi coefficient as our measure of association. Our estimated value of phi is .34. Phi indicates that there is a moderate association or correlation between type of institution and job satisfaction (remember that the phi coefficient is always positive).

2. a. The independent variable is whether or not the neighborhood is socially disorganized, and the dependent variable is the neighborhood crime rate.
   b. There are a total of 250 observations.
   c. There are 100 socially organized neighborhoods and 150 socially disorganized neighborhoods.
   d. There are 188 low crime rate neighborhoods and 62 high crime rate neighborhoods.
   e. This is a 2 × 2 contingency table.
   f. There are 52 socially disorganized neighborhoods with high crime rates.
   g. There are 90 socially organized neighborhoods with low crime rates.
   h. There is 1 degree of freedom.
   i. The risk of a high crime rate for those neighborhoods that are socially organized is .10, whereas for socially disorganized neighborhoods it is .35. This would suggest that social disorganization in the neighborhood is related to the crime rate.
   j. **Step 1:** H_0: Social organization in the neighborhood and neighborhood crime rates are independent.

      H_1: Social organization in the neighborhood and neighborhood crime rates are not independent.

   Step 2: Our test statistic is a chi-square test of independence, which has a chi-square distribution.

   Step 3: With 1 degree of freedom and an alpha of .01, our $\chi^2_{crit} = 21.85$. The critical region is any obtained chi-square to the right of this. Our decision rule is to reject the null hypothesis when $\chi^2_{obt} = 3.841$.

   Step 4: When we calculate our obtained chi-square, we find that it is $\chi^2_{obt} \geq 3.841$.

   Step 5: With a critical value of 6.635 and an obtained chi-square statistic of 20.07, our decision is to reject the null hypothesis. Our conclusion is that social organization in the neighborhood and neighborhood crime rates are not independent; there is a relationship between these two variables in the population.

We could use several different measures of association for a 2 × 2 contingency table. Our estimated value of Yule's *Q* would be .65, which would tell us that there is a strong positive relationship between social disorganization and neighborhood crime rate. More specifically, we would conclude that socially disorganized neighborhoods have higher rates of crime than socially organized neighborhoods. Since we have a 2 × 2 table, we could also have used the phi coefficient as our measure of association. Our estimated value of phi is .28. Phi indicates that there is a weak association or correlation between type of institution and job satisfaction (remember that the phi coefficient is always positive).

3. a. The independent variable is the jurisdiction where a defendant was tried, and the dependent variable is the type of sentence the defendant received.
   b. There are a total of 425 observations.
   c. There are 80 defendants from rural jurisdictions, 125 from suburban courts, and 220 who were tried in urban courts.
   d. There are 142 defendants who received jail time only, 95 who were fined and sent to jail, 112 who were sentenced to less than 60 days of jail time, and 76 who were sentenced to 60 or more days of jail.
   e. This is a 4 × 3 contingency table.
   f. There are 38 defendants from suburban courts who received less than 60 days of jail time as their sentence.
   g. There are 22 defendants tried in rural courts who received a sentence of a fine and jail.
   h. There are (4 – 1) × (3 – 1) or 6 degrees of freedom.
   i. The risk of 60 or more days of jail time is .20 for those tried in rural courts, .16 for those tried in suburban courts, and .18 for those tried in urban courts. There seems to be a slight relationship here, with those tried in rural courts more likely to be sentenced to more than 60 days of jail time.
   j. **Step 1:** H_0: Place where tried and type of sentence are independent.

      H_1: Place where tried and type of sentence are not independent.

      Step 2: Our test statistic is a chi-square test of independence, which has a chi-square distribution.

      Step 3: With 6 degrees of freedom and an alpha of .01 our $\chi^2_{\text{obt}} = .25$. The critical region is any obtained chi-square to the right of this. Our decision rule is to reject the null hypothesis when $\chi^2_{\text{obt}} \geq 16.812$.

      Step 4: When we calculate our obtained chi-square, we find that it is $\chi^2_{\text{crit}} = 21.85$.

      Step 5: With a critical value of 16.812 and an obtained chi-square statistic of 21.85, our decision is to reject the null hypothesis. Our conclusion is there is a relationship between where in the state a defendant was tried and the type of sentence the defendant received. Location of the trial and type of sentence are both nominal-level variables.

4. a. The independent variable is gender, and the dependent variable is the number of property crimes committed.
   b. There are a total of 360 cases.
   c. There are 257 persons who committed 0 to 4 property crimes and 103 persons who committed 5 or more property crimes.
   d. There are 110 non-White persons and 250 White persons.
   e. This is a 2 × 2 table.
   f. A total of 33 non-White offenders committed 5 or more property offenses.
   g. A total of 180 White offenders committed 0 to 4 property crimes.
   h. There is 1 degree of freedom.
   i. For non-Whites, the risk of 5 or more property crimes is .30, and for Whites it is .28. There does not seem to be much of a difference in the relative risks.
   j. **Step 1:** H_0: Race and the number of property crimes committed are independent.

      H_1: Race and the number of property crimes committed are not independent.

      Step 2: Our test statistic is a chi-square test of independence, which has a chi-square distribution.

Step 3: With 1 degree of freedom and an alpha of .05, our $\chi^2_{obt} = 3.841$. The critical region is any obtained chi-square to the right of this. Our decision rule is to reject the null hypothesis when $\chi^2_{obt} \geq 3.841$.

Step 4: When we calculate our obtained chi-square, we find that it is $\chi^2_{obt} = .25$.

Step 5: With a critical value of 3.841 and an obtained chi-square statistic of .25, our decision is to fail to reject the null hypothesis. Our conclusion is that one's race and the number of property crimes one commits are independent. There is no relationship in the population between these two variables. Since our conclusion is that there is no relationship in the population, we do not need to calculate a measure of association.

5. a. Employment is the independent variable, and the number of arrests within 3 years after release is the dependent variable.

   b. There are 115 observations or cases.

   c. There are 45 persons who reported having stable employment, 30 who reported having sporadic employment, and 40 who reported being unemployed.

   d. There are 54 persons who had no arrests within 3 years and 61 who had one or more arrests.

   e. This is a 2×3 contingency table.

   f. A total of 16 persons who were sporadically employed had one or more arrests.

   g. A total of 10 unemployed persons had no arrests.

   h. There are $(2 - 1) \times (3 - 1) = 2$ degrees of freedom.

   i. For those with stable employment, the risk of having one or more arrests is .33; for those with sporadic employment, it is .53; and for the unemployed, it is .75. The relative risk of at least one arrest increases as the individual's employment situation becomes worse.

   j. **Step 1:** H_0: Employment status and the number of arrests within 3 years are independent.

   H_1: Employment status and the number of arrests within 3 years are not independent.

   Step 2: Our test statistic is a chi-square test of independence, which has a chi-square distribution.

   Step 3: With 2 degrees of freedom and an alpha of .05, our $\chi^2_{obt} = 5.991$. The critical region is any obtained chi-square to the right of this. Our decision rule is to reject the null hypothesis when $\chi^2_{obt} \geq 5.991$.

   Step 4: When we calculate our obtained chi-square, we find that it is $\chi^2_{obt} = 15.33$.

   Step 5: With a critical value of 5.991 and an obtained chi-square statistic of 15.33, our decision is to reject the null hypothesis. Our conclusion is that employment status and the number of arrests after release are not independent. There is a relationship between the two variables in the population.

   Since both employment status and the number of arrests within 3 years are ordinal-level variables, we will use gamma as our measure of association. The value of gamma is

$$\gamma = \frac{1{,}800 - 520}{1{,}800 + 520}$$

$$\gamma = .55$$

   There is a moderately strong positive association between employment status and the number of arrests. More specifically, as one moves from stable employment to sporadic employment to unemployed, the risk of having one or more arrests increases.

6. a. The independent variable is whether an offender has tattoos.

   b. There are 320 total observations.

   c. There are 93 persons who committed 0 to 4 adult offenses, 78 who committed 5 to 9 adult offenses, 71 who committed 10 to 14 adult offenses, and 78 who committed 15 or more offenses as an adult.

d. There are 137 offenders with tattoos and 183 offenders without tattoos.

e. This is a 2×4 contingency table.

f. A total of 37 tattooed offenders have 10 to 14 adult offenses.

g. A total of 15 tattooed offenders have 15 or more adult offenses.

h. There are $(2 - 1)(4 - 1) = 3$ degrees of freedom.

i. The risk of 0 to 4 adult offenses is .43 for those without tattoos and .11 for those with tattoos. This would suggest that tattooed offenders commit more offenses as adults.

j. **Step 1:** H_0: Tattoo status and number of adult arrests are independent.

H_1: Tattoo status and number of adult arrests are not independent.

Step 2: Our test statistic is a chi-square test of independence, which has a chi-square distribution.

Step 3: With 3 degrees of freedom and an alpha of .01, our $\chi^2_{crit} = 11.345$. The critical region is any obtained chi-square to the right of this. Our decision rule is to reject the null hypothesis when $\chi^2_{obt} \geq 11.345$.

Step 4: When we calculate our obtained chi-square, we find that it is $\chi^2_{obt} = 84.92$.

Step 5: With a critical value of 11.345 and an obtained chi-square statistic of 84.92, our decision is to reject the null hypothesis. Our conclusion is that having tattoos and the number of adult arrests are not independent. There is a relationship between the two variables in the population.

The correct measure of association is a bit tricky. You could say that tattoo status is an ordinal-level variable since those with tattoos have more tattoos than those without tattoos. Since the number of adult offenses is also ordinal, you could use gamma. If you decided to do this, the value of gamma is

$$\gamma = \frac{17{,}258 - 3{,}208}{17{,}258 + 3{,}208}$$

$$\gamma = .69$$

There is a strong positive association between tattoo status and the number of adult offenses. More specifically, tattooed offenders have more adult offenses than those without tattoos.

2CHAPTER 9

1. An independent variable is the variable whose effect or influence on the dependent variable is what you want to measure. In causal terms, the independent variable is the cause, and the dependent variable is the effect. Low self-control is taken to affect one's involvement in crime, so self-control is the independent variable and involvement in crime is the dependent variable.

2. An independent-samples *t* test should be used whenever the two samples have been selected independently of one another. In an independent-samples *t* test, the sample elements are not related to one another. In a dependent-samples or matched-groups *t* test, by contrast, the sample elements are not independent but instead are related to one another. An example of dependent samples occurs when the same sample elements or persons are measured at two different points in time, as in a "before-and-after" experiment. A second common type of dependent sample is a matched-groups design.

3. The null and alternative hypotheses are

$$H_0: \mu = \mu_2$$

$$H_1: \mu_1 < \mu_2$$

The correct test is the pooled variance independent-samples *t* test, and our sampling distribution is Student's *t* distribution. We reject the null hypothesis if $t_{obt} \le -2.390$. The obtained value of *t* is

$$t_{obt} = \frac{2.1 - 8.2}{\sqrt{\frac{[(40-1)(1.8)^2] + [(25-1)(1.9)^2]}{40+25-2}}\sqrt{\frac{40+25}{(40)(25)}}}$$

$$t_{obt} = -13.01$$

Because our obtained value of *t* is less than the critical value and falls in the critical region, we decide to reject the null hypothesis of equal means. We conclude that those whose peers would disapprove of their driving drunk actually drive drunk less frequently than those whose coworkers are more tolerant of driving drunk.

4. The null and alternative hypotheses are

$$H_0: P_1 = P_2$$

$$H_1: P_1 > P_2$$

Because this problem involves two population proportions, our test statistic is the *z* test and our sampling distribution is the *z* or standard normal distribution. Our decision rule is to reject the null hypothesis if $z_{obt} \le -1.96$ or $z_{obt} \ge 1.96$.

The value of z_{obt} is

$$z_{obt} = \frac{.33 - .38}{\sqrt{(.35)(.65)}\sqrt{\frac{150+110}{(150)(110)}}}$$

$$z_{obt} = -.84$$

Because z_{obt} is not less than –1.96 or greater than 1.96 and does not fall in the critical region, we decide not to reject the null hypothesis. We cannot, therefore, reject the notion that the proportion who are rearrested is not different between those given fines and those given prison sentences.

5. The null and alternative hypotheses are

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 < \mu_2$$

The problem instructs you not to presume that the population standard deviations are equal ($\sigma_1 \ne \sigma_2$), so the correct statistical test is the separate variance *t* test, and the sampling distribution is Student's *t* distribution. With approximately 60 degrees of freedom and an alpha of .05 for a one-tailed test, the critical value of *t* is –1.671. The value for t_{obt} is

$$t_{obt} = \frac{18.8 - 21.3}{\sqrt{\frac{(4.5)^2}{50-1} + \frac{(3.0)^2}{25-1}}}$$

$$t_{obt} = -2.82$$

Because $t_{obt} \le t_{crit}$, we reject the null hypothesis of equal population means. Our conclusion is that the mean score on the domestic disturbance scale is significantly lower for males than for females. In other words, males are less likely to see the fair handling of domestic disturbances as an important part of police work.

6. The null and alternative hypotheses are

$$H_0: \mu_D = 0$$

$$H_1: \mu_D > 0$$

Our sample members (the judges) were deliberately matched in order to be comparable, so we have matched samples. The appropriate test statistic, then, is the dependent-samples or matched-groups *t* test, and the sampling distribution is Student's *t* distribution. Our decision rule is to reject the null hypothesis if $t_{obt} \geq 2.624$. The value of t_{obt} is

$$t_{obt} = \frac{1.4}{2.55/\sqrt{15-1}}$$

$$t_{obt} = 2.05$$

Because our t_{obt} is not greater than or equal to 2.624, we do not reject the null hypothesis. There is no difference in the number of capital cases lost on appeal between trained and untrained judges.

7. The null and alternative hypotheses are

$$H_0: P_1 = P_2$$

$$H_1: P_1 > P_2$$

Because this is a difference of proportions problem, the correct test statistic is the *z* test, and our sampling distribution is the *z* or standard normal distribution. Our decision rule is to reject the null hypothesis if $z_{obt} \geq 2.33$.

The value of z_{obt} is

$$z_{obt} = \frac{.43 - .17}{\sqrt{(.32)(.68)}\sqrt{\frac{100+75}{(100)(75)}}}$$

$$z_{obt} = 3.65$$

Because our obtained *z* is greater than the critical value of *z* (2.33) and z_{obt} falls in the critical region, we reject the null hypothesis. Delinquent children have a significantly higher proportion of criminal parents than do non-delinquent children.

8. The null and alternative hypotheses are

$$H_0: \mu_D = 0$$

$$H_1: \mu_D > 0$$

Because the two samples are the same youths at two points in time (before and after dropping out), we have dependent samples. The correct test statistic, then, is the dependent-samples or matched-groups *t* test, and the sampling distribution is Student's *t* distribution. Our decision rule is to reject the null hypothesis if $t_{obt} \leq -2.228$ or $t_{obt} \geq 2.228$.

The value of t_{obt} is

$$t_{\text{obt}} = \frac{-.27}{3.07/\sqrt{11}}$$
$$t_{\text{obt}} = -.29$$

Because our critical value of t is neither greater than or equal to 2.228 nor less than or equal to –2.228, and it does not fall in the critical region, we fail to reject the null hypothesis. We cannot reject the assumption that the number of delinquent offenses committed before dropping out is the same as the number after dropping out.

CHAPTER 10

1. An analysis of variance (ANOVA) can be performed whenever we have a continuous (interval- or ratio-level) dependent variable and a categorical variable with three or more levels or categories and we are interested in testing hypotheses about the equality of our population means.
2. If we have a continuous dependent variable and a categorical independent variable with only two categories or levels, the correct statistical test is a two-sample t test, assuming that the hypothesis test involves the equality of two population means.
3. It is called the analysis of variance because we make inferences about the differences among population means based on a comparison of the *variance* that exists within each sample relative to the variance that exists between the samples. More specifically, we examine the ratio of variance between the samples to the variance within the samples. The greater this ratio, the more between-samples variance there is relative to within-sample variance. Therefore, as this ratio becomes greater than 1, we are more inclined to believe that the samples were drawn from different populations with different population means.
4. As suggested in the answer to the last question, the two types of variance we use in the ANOVA F test are the variance between the samples and the variance within the samples:

$$F = \frac{\text{Between-groups variance}}{\text{Within-group variance}}$$

5. The formulas for the three degrees of freedom are

$$df_{\text{total}} = n - 1$$
$$df_{\text{between}} = k - 1$$
$$df_{\text{within}} = n - k$$

To check your arithmetic, make sure that $df_{\text{total}} = df_{\text{between}} + df_{\text{within}}$.

6. a. The independent variable is the level of stress reported by the women in the sample, and the dependent variable is the number of times they physically punished their children during the past month.

b. The total sum of squares is

$$SS_{\text{total}} = 343.2$$

The between-groups sum of squares is

$$SS_{\text{between}} = 242.6$$

The within-group sum of squares can be found by subtraction:

$$SS_{\text{within}} = 100.6$$

c.

$$df_{\text{between}} = k - 1 = 3 - 1 = 2$$

$$df_{\text{within}} = n - k = 30 - 3 = 27$$

$$df_{\text{total}} = n - 1 = 30 - 1 = 29$$

You can see that

$$df_{\text{between}} + df_{\text{within}} = df_{\text{total}}.$$

The ratio of sum of squares to degrees of freedom can now be determined:

$$SS_{\text{between}} / df_{\text{between}} = 242.6 / 2 = 121.30$$

$$SS_{\text{within}} / df_{\text{within}} = 100.6 / 27 = 3.73$$

The F ratio is

$$F_{\text{obt}} = 121.30 / 3.73 = 32.52.$$

d.

$$H_0\text{: } \mu_{\text{high stress}} = \mu_{\text{medium stress}} = \mu_{\text{low stress}}$$

$$H_1\text{: } \mu_{\text{high stress}} \neq \mu_{\text{medium stress}} \neq \mu_{\text{low stress}}$$

Our decision rule will be to reject the null hypothesis if $F_{\text{obt}} \geq 3.35$.

$F_{\text{obt}} = 32.52$. Since our obtained value of F is greater than the critical value, our decision is to reject the null hypothesis. We conclude that the population means are not equal and that the frequency of using physical punishment against one's child does vary with the amount of stress the woman feels.

e. Going to the studentized q table, you find the value of q to be equal to 3.49. To find the critical difference, you plug these values into your formula:

$$\text{CD} = 3.49\sqrt{\frac{3.73}{10}}$$

$$\text{CD} = 2.13$$

The critical difference for the mean comparisons, then, is 2.13. Find the absolute value of the difference between each pair of sample means in the problem to test each null hypothesis.

$$H_0: \mu_{\text{high stress}} = \mu_{\text{medium stress}}$$
$$H_1: \mu_{\text{high stress}} \neq \mu_{\text{medium stress}}$$

$$\begin{array}{lr} \text{High stress} & 8.30 \\ \text{Medium stress} & \underline{-3.30} \\ & |5.00| \end{array}$$

Since the absolute value of the difference between these sample means is greater than the critical difference score of 2.13, we would reject the null hypothesis. We would conclude that the frequency of using physical punishment for mothers experiencing high stress is greater than that for mothers experiencing medium stress.

$$H_0: \mu_{\text{high stress}} = \mu_{\text{low stress}}$$
$$H_1: \mu_{\text{high stress}} \neq \mu_{\text{low stress}}$$

$$\begin{array}{lr} \text{High stress} & 8.30 \\ \text{Low stress} & \underline{-1.60} \\ & |6.70| \end{array}$$

Because the absolute value of the difference between these sample means is greater than the critical difference score of 2.13, we would reject the null hypothesis and conclude that, on average, mothers under high stress use physical punishment more frequently than mothers under low stress.

$$H_0: \mu_{\text{medium stress}} = \mu_{\text{low stress}}$$
$$H_1: \mu_{\text{medium stress}} \neq \mu_{\text{low stress}}$$

$$\begin{array}{lr} \text{Medium stress} & 3.30 \\ \text{Low stress} & \underline{-1.60} \\ & |1.70| \end{array}$$

The absolute value of the difference between medium stress and low stress mothers is not greater than the critical difference score of 2.13. There is no significant difference between mothers experiencing medium stress and those experiencing low stress in the frequency with which they resort to physical punishment.

f. Eta squared is

$$\eta^2 = \frac{242.6}{343.2}$$
$$\eta^2 = .71$$

This tells us that there is a moderately strong relationship between a woman's feelings of stress and the frequency with which she uses physical punishment against her children. Specifically, about 71 percent of the variability in the frequency of physical punishment is explained by the mother's feelings of stress.

7. a. The independent variable is the state's general policy with respect to drunk driving ("get tough," make a "moral appeal," or not do much), and the dependent variable is the drunk driving rate in the state.

   b. The correct degrees of freedom for this table are

$$df_{\text{between}} = k - 1 = 3 - 1 = 2$$

$$df_{\text{within}} = n - k = 45 - 3 = 42$$

$$df_{\text{total}} = n - 1 = 45 - 1 = 44$$

You can see that

$$df_{\text{between}} + df_{\text{within}} = df_{\text{total}}.$$

The ratio of sum of squares to degrees of freedom can now be determined:

$$SS_{\text{between}}/df_{\text{between}} = 475.3/2 = 237.65$$

$$SS_{\text{within}}/df_{\text{within}} = 204.5/42 = 4.87$$

The F ratio is

$$F_{\text{obt}} = 237.65 / 4.87 = 48.80.$$

   c. The null and alternative hypotheses are

$$H_0: \mu_{\text{get tough}} = \mu_{\text{moral appeal}} = \mu_{\text{control}}$$

$$H_1: \mu_{\text{get tough}} \neq \mu_{\text{moral appeal}} \neq \mu_{\text{control}}$$

   Our decision rule will be to reject the null hypothesis if $F_{\text{obt}} \geq 5.18$.

   $F_{\text{obt}} = 48.80$. Since our obtained value of F is greater than the critical value, our decision is to reject the null hypothesis. We conclude that the population means are not equal.

   d. Going to the studentized table, you find the value of q to be equal to 4.37. To find the critical difference, you plug these values into your formula:

$$\text{CD}=4.37\sqrt{\frac{4.87}{15}}$$

$$\text{CD} = 2.49$$

The critical difference for the mean comparisons, then, is 2.49. Find the absolute value of the difference between each pair of sample means and test each null hypothesis:

$$H_0: \mu_{\text{get tough}} = \mu_{\text{moral appeal}}$$

$$H_1: \mu_{\text{get tough}} \neq \mu_{\text{moral appeal}}$$

$$\begin{array}{lr} \text{"Get Tough"} & 125.2 \\ \text{"Moral Appeal"} & -\underline{119.7} \\ & |5.5| \end{array}$$

Since the absolute value of the difference in sample means is greater than the critical difference score of 2.49, we would reject the null hypothesis. States that make a "moral appeal" have significantly lower levels of drunk driving on average than do states that "get tough."

$$H_0: \mu_{\text{get tough}} = \mu_{\text{control}}$$

$$H_1: \mu_{\text{get tough}} \neq \mu_{\text{control}}$$

$$\begin{array}{lr} \text{"Get Tough"} & 125.2 \\ \text{"Control"} & -\underline{145.3} \\ & |20.1| \end{array}$$

Since the absolute value of the difference in sample means is greater than the critical difference score of 2.49, we would reject the null hypothesis. States that "get tough" with drunk driving by increasing the penalties have significantly lower levels of drunk driving on average than do states that do nothing.

$$H_0: \mu_{\text{moral appeal}} = \mu_{\text{control}}$$

$$H_1: \mu_{\text{moral appeal}} \neq \mu_{\text{control}}$$

$$\begin{array}{lr} \text{"Moral Appeal"} & 119.7 \\ \text{"Control"} & -\underline{145.3} \\ & |25.6| \end{array}$$

The "moral appeal" states have significantly lower levels of drunk driving than do the "control" states. It appears, then, that doing *something* about drunk driving is better than doing little or nothing.

e. Eta squared is

$$\eta^2 = \frac{475.3}{679.8}$$

$$\eta^2 = .70$$

This tells us that there is a moderately strong relationship between the state's response to drunk driving and the rate of drunk driving in that state. Specifically, about 70% of the variability in levels of drunk driving is explained by the state's public policy.

8. a. The independent variable is the fear felt by people in a certain area of the city, and the dependent variable is the number of times a person was victimized during the past 5 years.

b. The correct degrees of freedom for this table are

$$df_{\text{between}} = k - 1 = 5 - 1 = 4$$

$$df_{\text{within}} = n - k = 250 - 5 = 245$$

$$df_{\text{total}} = n - 1 = 250 - 1 = 249$$

$$SS_{\text{between}} / df_{\text{between}} = 12.5 / 4 = 3.125$$

$$SS_{\text{within}} / df_{\text{within}} = 616.2 / 245 = 2.515$$

The F ratio is

$$F_{obt} = 3.125 / 2.515 = 1.24.$$

c.

$$H_0: \mu_{\text{very high}} = \mu_{\text{high}} = \mu_{\text{medium}} = \mu_{\text{low}} = \mu_{\text{very low}}$$

$$H_1: \mu_{\text{very high}} \neq \mu_{\text{high}} \neq \mu_{\text{medium}} \neq \mu_{\text{low}} \neq \mu_{\text{very low}}$$

Our decision rule will be to reject the null hypothesis if $F_{obt} \geq 2.37$.

$F_{obt} = 1.24$. Since our obtained value of F is not greater than or equal to the critical value, our decision is to not reject the null hypothesis. We conclude that different fear spots are not different in terms of their actual risk of victimization.

d. Since we failed to reject the null hypothesis, Tukey's HSD test is not appropriate.

e. The value of eta squared is

$$\eta^2 = \frac{12.5}{628.7}$$

$$\eta^2 = .02$$

There is no relationship between a person's fear of a given geographical area and the actual frequency of criminal victimization in that area. Only 2 percent of the variability in fear spots is explained by victimization levels.

9. a. The independent variable is the number of friends each girl has, and the dependent variable is the number of delinquent acts each girl is encouraged to commit.

b. The total sum of squares = 154

The between-groups sum of squares = 98

The within-group sum of square = 56

c. The correct degrees of freedom for this table are

$$df_{\text{between}} = k - 1 = 3 - 1 = 2$$

$$df_{\text{within}} = n - k = 21 - 3 = 18$$

$$df_{\text{total}} = n - 1 = 21 - 1 = 20$$

You can see that

$$df_{\text{between}} + df_{\text{within}} = df_{\text{total}}.$$

The ratio of sum of squares to degrees of freedom can now be determined:

$$SS_{\text{between}} / df_{\text{between}} = 98 / 2 = 49$$

$$SS_{\text{within}} / df_{\text{within}} = 56 / 18 = 3.11$$

The F ratio is

$$F_{obt} = \frac{49}{3.11} = 15.75$$

d. The null and alternative hypotheses are

$$H_0: \mu_{\text{a lot}} = \mu_{\text{some}} = \mu_{\text{a few}}$$

$$H_1: \mu_{\text{a lot}} \neq \mu_{\text{some}} \neq \mu_{\text{a few}}$$

With an alpha of .05 and 2 between-groups and 18 within-group degrees of freedom, our critical value of *F* is 3.55. Our decision rule is to reject the null hypothesis when $F_{obt} \geq 3.55$. The obtained *F* is 15.75; $F_{obt} > F_{crit}$, so our decision is to reject the null hypothesis and conclude that some of the population means are different from each other.

e. The value of the critical difference score is

$$CD = 3.61\sqrt{\frac{3.11}{7}}$$

$$CD = 2.41$$

A sample mean difference equal to or greater than an absolute value of 2.41 will lead us to reject the null hypothesis. We will now conduct a hypothesis test for each pair of population means.

$$H_0: \mu_{\text{a lot}} = \mu_{\text{some}}$$

$$H_1: \mu_{\text{a lot}} \neq \mu_{\text{some}}$$

"a lot"	7
"some"	−6
	|1|

Since this difference is less than the critical difference score of 2.41, we will fail to reject the null hypothesis. Girls who have a lot of friends are no different in the number of delinquent acts they are encouraged to commit than girls who have some friends.

$$H_0: \mu_{\text{a lot}} = \mu_{\text{a few}}$$

$$H_1: \mu_{\text{a lot}} \neq \mu_{\text{a few}}$$

"a lot"	7
"a few"	−2
	|5|

Since this difference is greater than the critical difference score of 2.41, we will reject the null hypothesis. Girls who have a lot of friends are encouraged to commit significantly more delinquent acts than girls who have a few friends.

$$H_0: \mu_{\text{some}} = \mu_{\text{a few}}$$

$$H_1: \mu_{\text{some}} \neq \mu_{\text{a few}}$$

"some"	6
"a few"	−2
	|4|

Since this difference is greater than the critical difference score of 2.41, we will reject the null hypothesis. Girls who have some friends are encouraged to commit significantly more delinquent acts than girls who have a few friends.

It would appear that Chapple and colleagues' (2014) hypothesis is correct. The presence of more friends in a friendship network puts females at higher risk of being encouraged to commit delinquent behavior.

f. Eta squared is

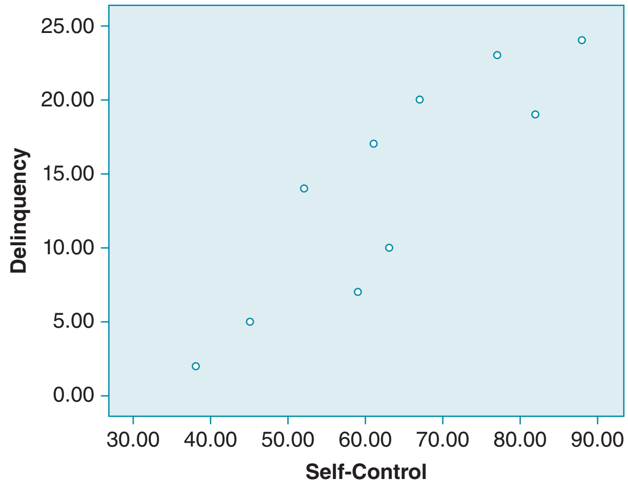
$$\eta^2 = \frac{98}{154}$$
$$\eta^2 = .64$$

There is a moderately strong relationship between the number of friends in a girl's friendship group and the number of delinquent acts she is encouraged to commit.

2CHAPTER 11

1. a. There is a moderate negative linear relationship between the median income level in a neighborhood and its rate of crime. As the median income level in a community increases, its rate of crime decreases.
   b. There is a weak positive linear relationship between the number of hours spent working after school and self-reported delinquency. As the number of hours spent working after school increases, the number of self-reported delinquent acts increases.
   c. There is a strong positive linear relationship between the number of prior arrests and the length of the current sentence. As the number of prior arrests increases, the length of the sentence received for the last offense increases.
   d. There is a weak negative linear relationship between the number of jobs held between the ages of 15 and 17 and the number of arrests as an adult.
   e. There is no linear relationship between a state's divorce rate and its rate of violent crime.
2. a. $(-.55)^2 = .30$, so 30% of the variation in neighborhood crime rates is explained by the median income level of the neighborhood.
   b. $(.17)^2 = .03$, so 3% of the variation in self-reported delinquency is explained by the number of hours a youth works after school.
   c. $(.74)^2 = .55$, so 55% of the variation in sentence length is explained by the number of prior arrests.
   d. $(-.12)^2 = .01$, so 1% of the variation in the number of arrests as an adult is explained by the number of jobs held when 15 to 17 years of age.
   e. $(-.03)^2 = .0009$, so less than 1% of the variation in a state's violent crime rate is explained by its divorce rate.
3. a. A $1 increase in the fine imposed decreases the number of price-fixing citations by .017.
   b. A 1% increase in unemployment increases the rate of property crime by .715.
   c. An increase of 1 year in education increases a police officer's salary by $1,444.53.

4 a. Scatterplot:



b. The value of the regression coefficient b is

$$b = \frac{10(9{,}890) - 632(141)}{10(42{,}230) - (632)^2}$$
$$b = \frac{98{,}900 - 89{,}112}{422{,}300 - 399{,}424}$$
$$b = .428$$

The value of the slope coefficient is .428. This tells us that an increase of 1 in an individual's score on the low self-control scale increases the number of self-reported criminal acts by .428.

c. The value of the y intercept is

$$14.1 = a + .428(63.2)$$
$$14.1 = a + 27.05$$
$$14.1 - 27.05 = a$$
$$a = -12.95$$

Thus, the value of the y intercept, or a, is equal to –12.95.

d. The predicted number of self-reported offenses when the score on the low self-control scale is equal to 70 can now be determined from our regression prediction equation:

$$\hat{y} = -12.95 + .428(70)$$
$$\hat{y} = -12.95 + 29.96$$
$$\hat{y} = -17.01$$

The predicted number of offenses, therefore, is 17.01.

e. We can use our computational formula to determine the value of r:

$$r = \frac{10(9{,}890) - 632(141)}{\sqrt{\left[10(42{,}230) - (632)^2\right]\left[10(2{,}529) - (141)^2\right]}}$$

$$r = \frac{9{,}788}{11{,}123.68}$$

$$r = .88$$

There is a strong positive correlation between low self-control and the number of self-reported criminal offenses.

We now want to conduct a hypothesis test about r. Our null hypothesis is that $r = 0$, and our alternative hypothesis is that $r > 0$. We predict direction because we have reason to believe that there is a positive correlation between low self-control and the number of self-reported crimes. To determine whether this estimated r value is significantly different from zero with an alpha level of .01, we calculate a t statistic with $n - 2$ degrees of freedom. The critical value of t with $10 - 2 = 8$ degrees of freedom, an alpha level of .01, and a one-tailed test is 2.896. Our decision rule is to reject the null hypothesis if $t_{obt} \geq 2.896$. Now, we calculate our t statistic:

$$t_{\text{obt}} = .88\sqrt{\frac{10-2}{1-(.88)^2}}$$

$$t_{\text{obt}} = .88(5.95)$$

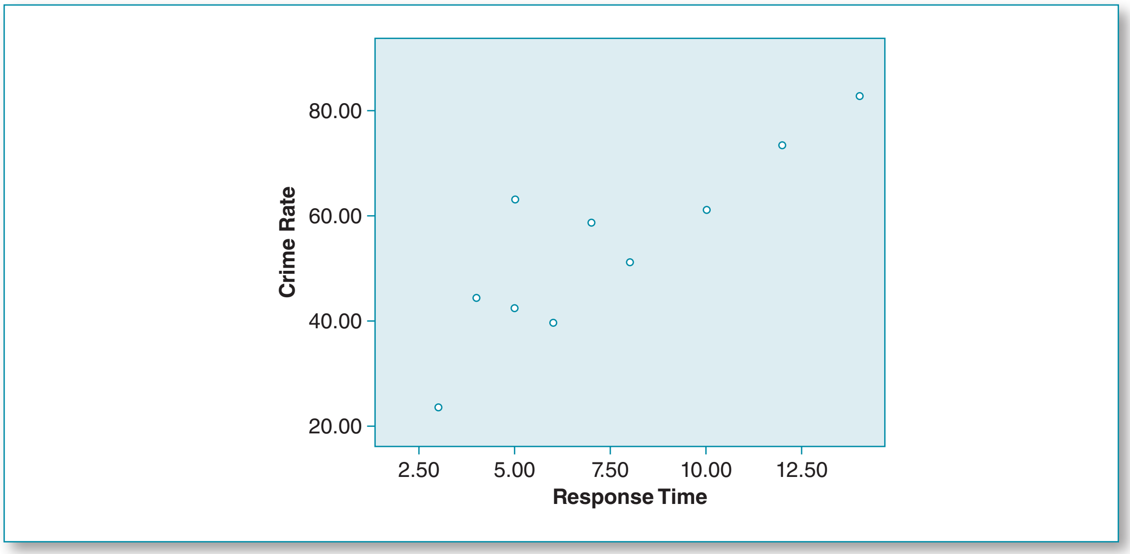
$$t_{\text{obt}} = 5.24$$

We have a t_{obt} of 5.24. Since $5.24 > 2.896$, we decide to reject the null hypothesis. There is a significant positive correlation between low self-control and self-reported crime in the population. As levels of self-control decrease, rates of self-reported crime increase.

f. Our r was .88, and $(.88)^2 = .77$, so 77% of the variation in self-reported crime is explained by low self-control.

g. Based on our results, we would conclude that there is a significant positive linear relationship between low self-control and self-reported offending. Our results are consistent with the findings of Burt and colleagues (2014) and with the theory of low self-control developed by Gottfredson and Hirschi (1990).

5. a. Scatterplot:



b. The value of the regression coefficient is

$$b = \frac{10(4{,}491.4) - 74(541.1)}{10(664) - (74)^2}$$

$$b = \frac{4{,}872.6}{1{,}164}$$

$$b = 4.19$$

The value of the slope coefficient is 4.19. This tells us that a 1-minute increase in police response time increases the crime rate by 4.19 per 1,000. The longer the response time, the higher the crime rate. Stated conversely, the shorter the response time, the lower the crime rate.

c. The value of the *y* intercept is

$$54.11 = a + 4.19(7.4)$$

$$54.11 = a + 31.01$$

$$54.11 - 31.01 = a$$

$$a = 23.1$$

Thus, the value of the *y* intercept, or *a*, is equal to 23.1.

d. The predicted community rate of crime when the police response time is 11 minutes can now be determined from our regression prediction equation:

$$\hat{y} = 23.1 + 4.19(11)$$

$$\hat{y} = 23.1 + 46.09$$

$$\hat{y} = 69.19$$

The predicted crime rate, therefore, is 69.19 crimes per 1,000 population.

e. The value of *r* is

$$r = \frac{10(4{,}491.4) - 74(541.1)}{\sqrt{\left[10(664) - (74)^2\right]\left[10(32{,}011.3) - (541.1)^2\right]}}$$

$$r = \frac{4{,}872.6}{5{,}639.6}$$

$$r = .86$$

There is a strong positive correlation between the time it takes the police to respond and the crime rate.

We now want to conduct a hypothesis test about *r*. Our null hypothesis is that $r = 0$, and our alternative hypothesis is that $r > 0$. We predict direction because we have reason to believe that there is a positive correlation between the number of minutes it takes the police to respond and the community's rate of crime (the longer the response time, the higher the crime rate). To determine whether this estimated *r* value is significantly different from zero with an alpha level of .05, we calculate a *t* statistic with $n - 2$ degrees of freedom. We go to the *t* table to find our critical value of t with 10 – 2 = 8 degrees of freedom, an alpha level of .05, and a one-tailed test. The critical value of t is 1.86. Our decision rule is to reject the null hypothesis if $t_{obt} \geq 1.86$. Now we calculate our t_{obt}:

$$t = .86\sqrt{\frac{10-2}{1-(.86)^2}}$$

$$t = .86(5.54)$$

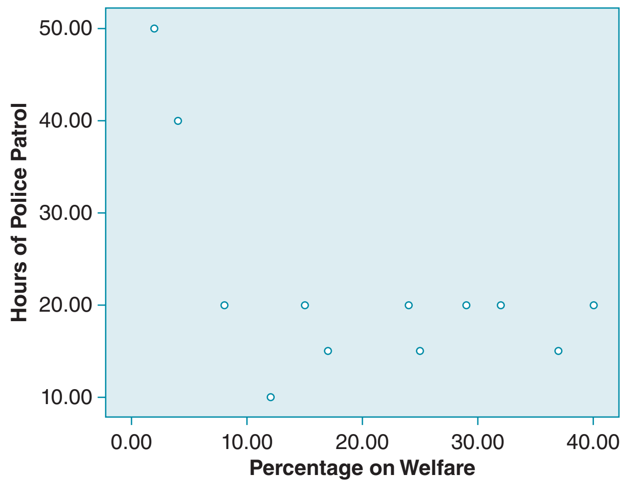
$$t = 4.76$$

We have a t_{obt} of 4.76. Since 4.76 > 1.86, we decide to reject the null hypothesis. There is a significant positive correlation between the length of police response time and community crime rate.

f. Our r was .86; $(.86)^2 = .74$, so 74% of the variation in community crime rates is explained by police response time.

g. Based on our results, we would conclude that there is a significant positive linear relationship between police response time and community crime rate.

6. a. Scatterplot:



b. The value of the slope coefficient is

$$b = \frac{12(4{,}525) - 245(265)}{12(6{,}777) - (245)^2}$$

$$b = \frac{54{,}300 - 64{,}925}{81{,}324 - 60{,}025}$$

$$b = -.499$$

The value of the slope coefficient is –.499. This tells us that a 1% increase in the percentage of the population that is on welfare decreases the hours of daily police patrol by .499 (about one half hour). The greater the percentage of the population receiving public assistance in a neighborhood, the lower the number of hours of police patrol.

c. The value of the *y* intercept is

$$22.1 = a + -.499(20.4)$$
$$22.1 = a + -10.18$$
$$22.1 + 10.18 = a$$
$$a = 32.28$$

Thus, the value of the *y* intercept, or *a,* is equal to 32.28.

d. The predicted number of daily hours of police patrol when the percentage receiving public assistance in the neighborhood is 30% can now be determined from our regression prediction equation:

$$\hat{y} = 32.28 + (-.499)(30)$$
$$\hat{y} = 32.28 - 14.97$$
$$\hat{y} = 17.31$$

The predicted number of hours of police patrol, therefore, is 17.31 hours.

e. From our calculations in part b above, we have all the information we need to solve for *r:*

$$r = \frac{12(4{,}525) - 245(265)}{\sqrt{\left[12(6{,}777) - (245)^2\right]\left[12(7{,}725) - (265)^2\right]}}$$
$$r = \frac{-10{,}625}{\sqrt{478{,}695{,}025}}$$
$$r = -.49$$

There is a moderately strong negative correlation between the percentage of the community on welfare and the number of daily hours of police patrol.

We now want to conduct a hypothesis test about *r.* Our null hypothesis is that $r = 0$, and our alternative hypothesis is that $r < 0$. We predict direction because we have reason to believe that there is a negative correlation between the affluence of the neighborhood and the number of hours of daily police patrol. To determine whether this estimated *r* value is significantly different from zero with an alpha level of .05, we calculate a *t* statistic with $n - 2$ degrees of freedom. We go to the *t* table to find our critical value of *t* with $12 - 2 = 10$ degrees of freedom, an alpha level of .05, and a one-tailed test. The critical value of *t* is –1.812. The critical value is negative because in our alternative hypothesis, we predicted that the population value of *r* was less than zero. Our decision rule is to reject the null hypothesis if $t_{obt} < -1.812$. Our obtained *t* value is

$$t = -.49\sqrt{\frac{12-2}{1-(-.49)^2}}$$
$$t = -.49(3.63)$$
$$t = -1.78$$

We have a t_{obt} of –1.78. Because –1.78 is not less than the critical value of –1.812, we fail to reject the null hypothesis. There is no significant correlation between the percentage of a neighborhood that is receiving public assistance and the number of hours of daily police patrol.

f. Our *r* was –.49, and $(-.49)^2 = .24$, so 24% of the variation in the number of hours of police patrol is explained by the percentage of the neighborhood that is receiving public assistance.

g. Based on our results, we would conclude that there is no significant linear relationship between the percentage of a neighborhood on welfare and the number of hours per day that the police patrol the neighborhood.

h. The values of b and r without community numbers 11 and 12 are

$$b = \frac{10(4{,}265) - 239(175)}{10(6{,}757) - (239)^2}$$

$$b = \frac{42{,}650 - 41{,}825}{67{,}570 - 57{,}121}$$

$$b = .079$$

$$r = \frac{10(4{,}265) - 239(175)}{\sqrt{\left[10(6{,}757) - (239)^2\right]\left[10(3{,}175) - (175)^2\right]}}$$

$$r = \frac{825}{3{,}428.6}$$

$$r = .24$$

We would conclude from these new data that there is not a very strong relationship between the percentage of neighborhood families on welfare and the number of hours of police patrol. You can see that without the last two observations, the slope of the data points changes. These last two data points are very unusual. They are unusually affluent neighborhoods, with 4% and 2% of the residents receiving public assistance, respectively. Moreover, they receive an unusually high number of hours of police patrols. These two neighborhoods, then, are outliers, and as outliers they can distort your data.

2CHAPTER 12

1. a. The least-squares regression equation for this problem is

$$y = a + b_1x_1 + b_2x_2$$

$$y = 19.642 + (.871)(\text{divorce rate}) + (-.146)(\text{age})$$

b. The partial slope coefficient for the variable "divorce" indicates that as the divorce rate per 100,000 population increases by 1, the rate of violent crime per 100,000 increases by .871 when controlling for the mean age of the state's population. The partial slope coefficient for the variable "age" indicates that as the mean age of the state's population increases by 1 year, the rate of violent crime per 100,000 decreases by .146 when controlling for the divorce rate. The intercept is equal to 19.642. This tells us that when both the divorce rate and the mean age are equal to zero, the rate of violent crime is 19.642 per 100,000.

c. The standardized regression coefficient for divorce is .594, and that for age is –.133. Based on this, we would conclude that the divorce rate is more influential in explaining state violent crime rates than is the mean age of the population. A second way to look at the relative strength of the independent variables is to compare the absolute values of their respective t ratios. The t ratio for divorce is 4.268, and that for age is –3.110. Based on this, we would conclude that the divorce rate is more influential in explaining rates of violence than is the mean age of a state's population.

d. The divorce rate and mean age together explain approximately 63% of the variance in rates of violent crime. The adjusted R^2 value is .61, indicating that 61% of the variance is explained.

e. The null and alternative hypotheses are

$$\mathrm{H}_0: \beta_1, \beta_2 = 0; \text{ or } R^2 = 0$$
$$\mathrm{H}_1: \beta_1, \text{or } \beta_2 \neq 0; \text{ or } R^2 \neq 0$$

$F_{\text{obt}} = 27.531$. The probability of obtaining an F of 27.531 if the null hypothesis were true is .00001. Since this probability is less than our alpha of .01, our decision is to reject the null hypothesis that all the slope coefficients are equal to zero.

$$\mathrm{H}_0: \beta_{\text{divorce}} = 0$$
$$\mathrm{H}_1: \beta_{\text{divorce}} \neq 0$$

$t_{\text{obt}} = 4.268$. The probability of obtaining a t this size if the null hypothesis were true is equal to .0001. Since this probability is less than our alpha level of .01, we decide to reject the null hypothesis. We conclude that the population partial slope coefficient for the effect of the divorce rate on the rate of violent crime is significantly different from 0.

$$\mathrm{H}_0: \beta_{\text{age}} = 0$$
$$\mathrm{H}_1: \beta_{\text{age}} \neq 0$$

$t_{\text{obt}} = -3.110$. The probability of obtaining a t statistic this low if the null hypothesis is true is equal to .0011. Since this is less than .01, we will reject the null hypothesis that there is no relationship in the population between the mean age of a state and its rate of violent crime.

2. a.

$$b_{\text{morale}} = \left(\frac{4.49}{3.47}\right)\left(\frac{(-.77)-(-.63)(.67)}{1-(.67)^2}\right)$$
$$b_{\text{morale}} = 1.29\left(\frac{(-.77)-(-.42)}{1-.45}\right)$$
$$b_{\text{morale}} = 1.29\left(\frac{-.35}{.55}\right)$$
$$b_{\text{morale}} = -.821$$

$$b_{\text{staff}} = \left(\frac{4.49}{.15}\right)\left(\frac{(-.63)-(-.77)(.67)}{1-(.67)^2}\right)$$
$$b_{\text{staff}} = 29.93\left(\frac{(-.63)-(-.52)}{1-.45}\right)$$
$$b_{\text{staff}} = 29.93\left(\frac{-.11}{.55}\right)$$
$$b_{\text{staff}} = -5.986$$

The partial slope coefficient for employee morale is –.821. This tells us that as the score on our measure of employee morale increases by 1, the number of jail escapes decreases by .821 (or about 1 less escape) when controlling for the staff-to-inmate ratio. The partial slope coefficient for the staff-to-inmate ratio is –5.986. This indicates that as the staff-to-inmate ratio increases by 1 unit, the number of jail escapes decreases by approximately 6.

b.

$$\begin{aligned} \bar{y} &= a + b_1\bar{x}_1 + b_2\bar{x}_2 \\ 9.43 &= a + (-.821)(6.07) + (-5.986)(.36) \\ 9.43 &= a - 7.13 \\ 9.43 + 7.13 &= a \\ 16.56 &= a \end{aligned}$$

The value of the intercept is 16.56, and the full regression equation is

$$y = 16.560 + (-.821)x_1 + (-5.986)\,x_2.$$

c.

$$\begin{aligned} \hat{y} &= 16.560 + (-.821)(8) + (-5.986)(.3) \\ \hat{y} &= 16.580 + (-6.568) + (-1.796) \\ \hat{y} &= 16.560 + (-8.364) \\ \hat{y} &= 8.196 \end{aligned}$$

With a staff morale score of 8 and a staff-to-inmate ratio of .3, we would predict that there would be approximately 8 escapes per year.

d. The beta weights are

$$\begin{aligned} b^*_{\text{morale}} &= (-.821)\left(\frac{3.47}{4.49}\right) \\ &= -.634 \\ b^*_{\text{staff}} &= (-5.986)\left(\frac{.15}{4.49}\right) \\ &= -.200 \end{aligned}$$

The beta weight for staff morale is –.634, and the beta weight for staff/inmate ratio is –.200. Because the beta weight for morale is greater than that for the staff-to-inmate ratio, it has a stronger effect on the number of jail escapes.

e.

$$\begin{aligned} R^2 &= (-.77)^2 + (-.245)^2\left(1 - (-.77)^2\right) \\ R^2 &= .59 + (.06)(.41) \\ R^2 &= .59 + .02 \\ R^2 &= .61 \end{aligned}$$

Together, staff morale and the staff-to-inmate ratio explain approximately 61 percent of the variance in jail escapes.

3. a. The least-squares regression equation from the supplied output would be

$$\breve{y} = 16.245 + (-1.467)(\text{ENV}) + (1.076)(\text{REL})$$

b. The partial slope coefficient for the environmental factors variable is –1.467. This tells us that as a person's score on the environmental causes of crime scale increases by 1, his or her score on the punitiveness scale decreases by 1.467 when controlling for religious fundamentalism. The partial slope coefficient for the religious fundamentalism scale is 1.076. As a person's score on religious fundamentalism increases by 1 unit, his or her score on the punitiveness scale increases by 1.076 when controlling for score on the environmental scale. The value of the intercept is 16.245. When both independent variables are zero, a person's score on the punitiveness scale is 16.245.

c. The beta weight for ENV is –.609. The beta weight for REL is .346. A 1-unit change in the ENV variable produces almost twice as much change in the dependent variable as the REL variable. Comparing these beta weights would lead us to conclude that the environmental factors scale is more important in explaining punitiveness scores than is the religious fundamentalism variable.

   The t ratio for ENV is –3.312, whereas that for REL is only 1.884. We would again conclude that ENV has the greater influence on the dependent variable.

d. The adjusted R^2 coefficient indicates that the environmental factors and religious fundamentalism scales together explain approximately 60% of the variance in the punitiveness measure.

e. The null and alternative hypotheses are

$$\text{H}_0 : \beta_{\text{ENV}}, \beta_{\text{REL}} = 0; \text{ or } R^2 = 0$$
$$\text{H}_1 : \beta_{\text{ENV}}, \text{or}\, \beta_{\text{REL}} \neq 0; \text{ or } R^2 \neq 0$$

   The probability of an F of 11.59 if the null hypothesis were true is .0016. Since this probability is less than our chosen alpha of .01, our decision is to reject the null hypothesis that all the slope coefficients are equal to zero.

$$\text{H}_0 : \beta_{\text{ENV}} = 0$$
$$\text{H}_1 : \beta_{\text{ENV}} < 0$$

   The output gives you a t_{obt} of –3.312, and the probability of obtaining a t this size if the null hypothesis were true is equal to .0062. Since this probability is less than our chosen alpha level of .01, we decide to reject the null hypothesis. We conclude that the population partial slope coefficient is less than 0.

$$\text{H}_0 : \beta_{\text{REL}} = 0$$
$$\text{H}_1 : \beta_{\text{REL}} > 0$$

   The t ratio for the variable ENV is $t_{\text{obt}} = 1.884$. The probability of obtaining a t statistic of this magnitude if the null hypothesis were true is equal to .0840. Since this probability is greater than .01, our decision is to fail to reject the null hypothesis. The partial slope coefficient between religious fundamentalism and punitiveness toward criminal offenders is not significantly different from zero in the population once we control for a belief in environmental causes of crime.

$$\hat{y} = 16.245 + (-1.467)(\text{ENV}) + (1.076)(\text{REL})$$
$$\hat{y} = 16.245 + (-1.467)(2) + (1.076)(8)$$
$$\hat{y} = 16.245 + (-2.934) + 8.608$$
$$\hat{y} = 21.919$$