

# Appendix C

## Solutions to Odd-Numbered Practice Problems

### 2CHAPTER 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.
3.
  - a. Quantitative; interval/ratio.
  - b. Quantitative; interval/ratio.
  - c. Quantitative; interval/ratio.
  - d. Qualitative; nominal.
  - e. Qualitative; nominal.
  - f. Quantitative; interval/ratio.
5. Arrest is the independent variable, and future drunk driving behavior 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.

9.

	<i>f</i>	<i>Proportion</i>	<i>%</i>
Less than \$10	16	.029	2.9
\$10–\$49	39	.072	7.2
\$50–\$99	48	.088	8.8
\$100–\$249	86	.159	15.9
\$250–\$999	102	.188	18.8
\$1,000 or more	251	.463	46.3
	<i>n</i> = 542		

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

### 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.

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	<i>f</i>	<i>p</i>	%
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.”

5. a.

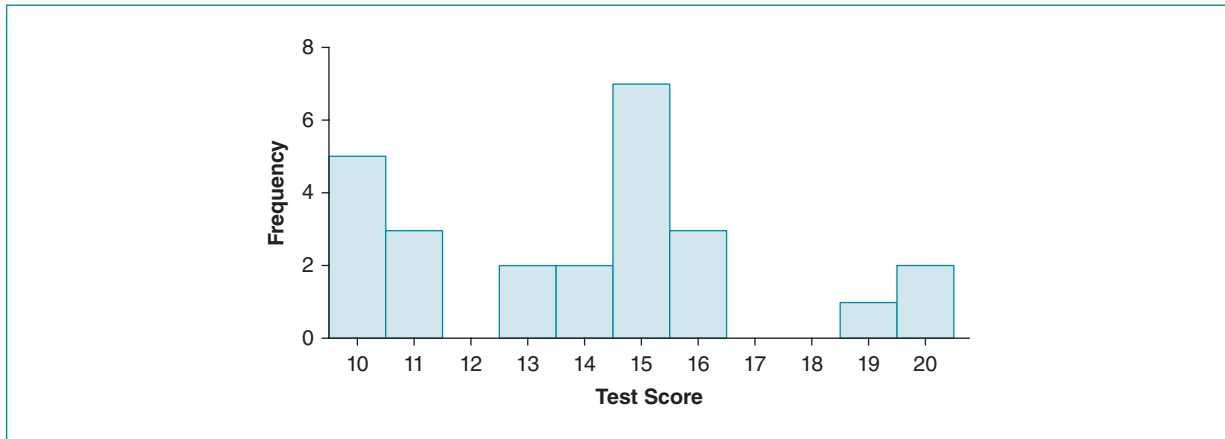
Value	<i>f</i>	<i>cf</i>	<i>p</i>	<i>cp</i>	%	<i>c%</i>
10	5	5	.20	.20	20	20
11	3	8	.12	.32	12	32
12	0	8	.00	.32	0	32
13	2	10	.08	.40	8	40
14	2	12	.08	.48	8	48
15	7	19	.28	.76	28	76
16	3	22	.12	.88	12	88
17	0	22	.00	.88	0	88
18	0	22	.00	.88	0	88
19	1	23	.04	.92	4	92
20	2	25	.08	1.00	8	100

b.

Value	<i>f</i>	<i>p</i>	%
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.”

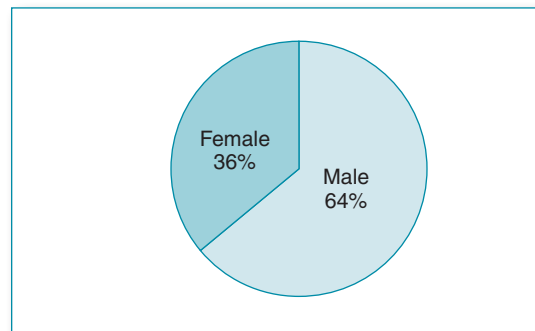
### Distribution of Test Scores for Recruit Class



- 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.

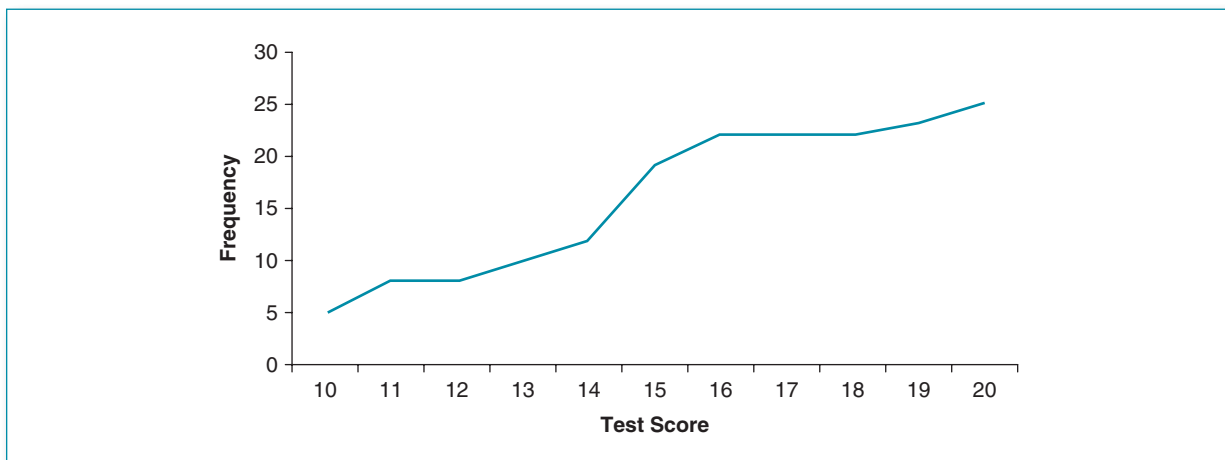
A pie chart of the percentages for the gender data is at right.

### Gender Distribution of Recruit Class

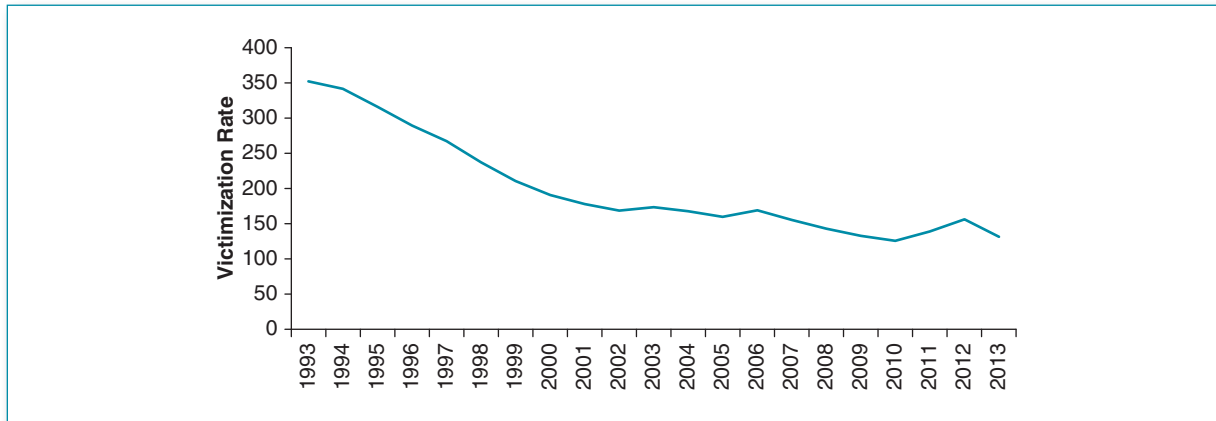


- h. A cumulative frequency distribution for the test scores would look like this:

### Cumulative Frequency Line Graph for Test Score Data



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



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 the time plot above.

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.

## CHAPTER 3

- 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.
- 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.

- 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.”
- Mean number of executions:

$$\bar{X} = \frac{(42 + 37 + 52 + 46 + 43 + 43 + 39 + 35)}{8}$$

$$\bar{X} = 42.125 \text{ executions per year}$$

<align at = sign>

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.

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

- 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.
- 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
				$\Sigma = 9,868.80$

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

$$s^2 = 48.38$$

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.

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.

## CHAPTER 5

- $P(x = \$30,000) = 16/110 = .145$
  - $P(x = \$35,000) = 7/110 = .064$
  - 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.
  - $P(x \geq \$31,000) = (19/110) + (12/110) + (15/110) + (8/110) + (7/110) = (61/110) = .555$

- e. There are two ways to calculate this probability. First:  
 $P(x \leq \$30,000) = (16/110) + (10/110) + (9/110) + (8/110) + (6/110) = (49/110) = .445$ . 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.  $P(x = \$28,000 \text{ or } \$30,000 \text{ or } \$31,500 \text{ or } \$32,000 \text{ or } \$32,500) = (10/110) + (16/110) + (19/110) + (12/110) + (15/110) = (72/110) = .655$
- g.  $P(x < \$25,000) = 0$
- h.  $P(x = \$28,000 \text{ or } \$32,000 \text{ or } \$35,000) = (10/110) + (12/110) + (7/110) = (29/110) = .264$
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.
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.
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.
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%.

## 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.

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.

5. The standard deviation of the sampling distribution is the standard deviation of an infinite number of sample estimates [means ( $X$ ) 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.
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:

$$95\% \text{ c.i.} = 560 \pm 2.145 \left( \frac{45}{\sqrt{15-1}} \right)$$

$$95\% \text{ c.i.} = 560 \pm 2.145 \left( \frac{45}{3.74} \right)$$

$$95\% \text{ c.i.} = 560 \pm 2.145(12.03)$$

$$95\% \text{ c.i.} = 560 \pm 25.8$$

$$534.2 \leq \mu \leq 585.8$$

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).

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.

## CHAPTER 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 ( $\sigma$ ) 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.
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  $z_{\text{obt}} > 2.33$ ). The value of  $z_{\text{obt}}$  is

$$z_{\text{obt}} = \frac{6.3 - 4.6}{1.9/\sqrt{64}}$$

$$z_{\text{obt}} = \frac{1.7}{.2375}$$

$$z_{\text{obt}} = 7.16$$

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.

5. The null and alternative hypotheses are

$$H_0: \mu = 25.9$$

$$H_1: \mu \neq 25.9$$

The critical value of  $z$  is  $\pm 2.58$ . The value of  $z_{\text{obt}}$  is

$$z_{\text{obt}} = \frac{27.3 - 25.9}{6.5/\sqrt{175}}$$

$$z_{\text{obt}} = \frac{1.40}{6.5/13.23}$$

$$z_{\text{obt}} = \frac{1.40}{.49}$$

$$z_{\text{obt}} = 2.86$$

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.

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  $t_{\text{obt}}$  is

$$t_{\text{obt}} = \frac{6.3 - 4}{1.5/\sqrt{12}}$$

$$t_{\text{obt}} = \frac{2.30}{1.5/3.46}$$

$$t_{\text{obt}} = \frac{2.30}{.43}$$

$$t_{\text{obt}} = 5.35$$

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.

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  $z_{\text{obt}}$  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%.

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  $z_{\text{obt}}$  is

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

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  $2 \times 2$  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  $\chi^2_{crit} = 3.841$ . 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  $\chi^2_{obt} = 21.11$ .

**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  $2 \times 2$  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  $2 \times 2$  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).

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  $4 \times 3$  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  $(4 - 1) \times (3 - 1)$  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  $\chi^2_{\text{crit}} = 16.812$ . The critical region is any obtained chi-square to the right of this. Our decision rule is to reject the null hypothesis when  $\chi^2_{\text{obt}} \geq 16.812$ .

**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.

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  $2 \times 3$  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  $\chi^2_{\text{obt}} = 5.991$ . The critical region is any obtained chi-square to the right of this. Our decision rule is to reject the null hypothesis when  $\chi^2_{\text{obt}} \geq 5.991$ .

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

**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.

## CHAPTER 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.
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  $t_{\text{obt}} \leq -2.390$ . The obtained value of  $t$  is

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

$$t_{\text{obt}} = -13.01$$

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.

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 ( $\sigma_1 \neq \sigma_2$ ), 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  $t_{\text{obt}}$  is

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

$$t_{\text{obt}} = -2.82$$

Because  $t_{\text{obt}} \leq t_{\text{crit}}$ , 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.

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  $z_{\text{obt}} \geq 2.33$ .

The value of  $z_{\text{obt}}$  is

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

$$z_{\text{obt}} = 3.65$$

Because our obtained  $z$  is greater than the critical value of  $z$  (2.33) and  $z_{\text{obt}}$  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.

## CHAPTER 10

1. An analysis of variance 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.
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.
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  $df_{\text{total}} = df_{\text{between}} + df_{\text{within}}$ .

- 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:

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

$$CD = 2.49$$

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:

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

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

“Get Tough”	125.2
-------------	-------

“Moral Appeal”	-119.7
	5.5

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.”

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

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

“Get Tough”	125.2
“Control”	-145.3
	20.1

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.

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

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

“Moral Appeal”	119.7
----------------	-------

“Control”	-145.3
	25.6

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.

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

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

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  $F_{\text{obt}} \geq 3.55$ . The obtained  $F$  is 15.75;  $F_{\text{obt}} > F_{\text{crit}}$ , 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”	<u>-6</u>
--------	-----------

	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”	<u>-2</u>
---------	-----------

	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”	<u>-2</u>
---------	-----------

	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.

## CHAPTER 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.
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.
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

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

$$54.11 = a + 31.01$$

$$54.11 - 31.01 = a$$

$$a = 23.1$$

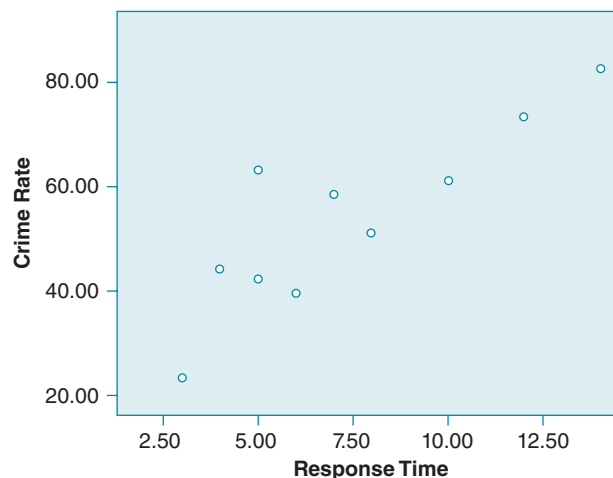
Thus, the value of the  $y$  intercept, or  $a$ , is equal to 23.1.

- 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

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

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

$$r = .86$$

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  $t_{\text{obt}} \geq 1.86$ . Now we calculate our  $t_{\text{obt}}$ :

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

$$t = .86(5.54)$$

$$t = 4.76$$

We have a  $t_{\text{obt}}$  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.

## CHAPTER 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

$$H_0: \beta_1, \beta_2 = 0; \text{ or } R^2 = 0$$

$$H_1: \beta_1, \text{ or } \beta_2 \neq 0; \text{ or } R^2 \neq 0$$

$F_{\text{obt}} = 27.531$ . 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.

$$H_0: \beta_{\text{divorce}} = 0$$

$$H_1: \beta_{\text{divorce}} \neq 0$$



$t_{\text{obt}} = 4.268$ . 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.

$$H_0: \beta_{\text{age}} = 0$$

$$H_1: \beta_{\text{age}} \neq 0$$

$t_{\text{obt}} = -3.110$ . 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.

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

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

- 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

$$H_0: \beta_{\text{ENV}}, \beta_{\text{REL}} = 0; \text{ or } R^2 = 0$$

$$H_1: \beta_{\text{ENV}}, \text{ or } \beta_{\text{REL}} \neq 0; \text{ or } R^2 \neq 0$$

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.

$$H_0: \beta_{\text{ENV}} = 0$$

$$H_1: \beta_{\text{ENV}} < 0$$

The output gives you a  $t_{\text{obt}}$  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.

$$H_0: \beta_{\text{REL}} = 0$$

$$H_1: \beta_{\text{REL}} > 0$$

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.

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

$$\hat{y} = 16.245 + (-1.467)(2) + (1.076)(8)$$

$$\hat{y} = 16.245 + (-2.934) + 8.608$$

$$\hat{y} = 21.919$$