ANSWERS TO ODD-NUMBERED EXERCISES

CHAPTER 1

1. Once our research question, the hypothesis, and the study variables have been selected, we move on to the next stage of the research process—measuring and collecting the data. The choice of a particular data collection method or instrument depends on our study objective. After our data have been collected, we have to find a systematic way to organize and analyze our data and set up some set of procedures to decide what we mean.

3.
   a. Interval-ratio
   b. Interval-ratio
   c. Nominal
   d. Ordinal
   e. Nominal
   f. Interval-ratio
   g. Ordinal

5. There are many possible variables from which to choose. Some of the most common selections by students will probably be as follows: type of occupation or industry, work experience, and educational training or expertise. Students should first address the relationship between these variables and gender. Students may also consider measuring structural bias or discrimination.

7.
   a. Annual income
   b. Gender—nominal; Number of hours worked per week—interval-ratio; Years of education—interval-ratio; Job title—nominal.
   c. This is an application of inferential statistics. She is using information based on her sample to predict the annual income of a larger population of young graduates.

9.
   a. Individual age: This variable could be measured as an interval-ratio variable, with actual age in years reported. As discussed in the chapter, interval-ratio variables are the highest level of measurement and can also be measured at ordinal or nominal levels.
   b. Annual income: This variable could be measured as an interval-ratio variable, with actual dollar earnings reported.

c. Religiosity: This variable could be measured in several ways. For example, as church attendance, the variable could be ordinal (number of times attended church in a month: every week, at least twice a month, less than two times a month, none at all).

d. Student performance: This could be measured as an interval-ratio variable as GPA or test score.

e. Social class: This variable is an ordinal variable, with categories low, working, middle, and upper.

f. Number of children: This variable could be measured in several ways. As an interval-ratio measure, the actual number of children could be reported. As an ordinal measure, the number of children could be measured in categories: 0, 1–2, 3–4, 5 or more. This could also be a nominal measurement—do you have children? Yes or No.

CHAPTER 2

1.

a. Race is a nominal variable. Class is an ordinal variable, since the categories can be ordered. Trauma is an interval variable.

b. Frequency Table for Race

Race	Frequency (f)
White	17
Nonwhite	13
Total (*N*)	30

c. White: 17/30 = .57%; Nonwhite: 13/30 = .43

3.

Number of Traumas	Frequency (f)
0	15
1	11
2	4
Total (*N*)	30

a. Trauma is an interval or ratio-level variable, since it has a real zero point and a meaningful numeric scale.

b. People in this survey are more likely to have experienced no traumas last year (50% of the group).

c. The proportion who experienced one or more traumas is calculated by first adding 36.7% and 13.3% = 50%. Then, divide that number by 100 to obtain the proportion, 0.50, or half the group.

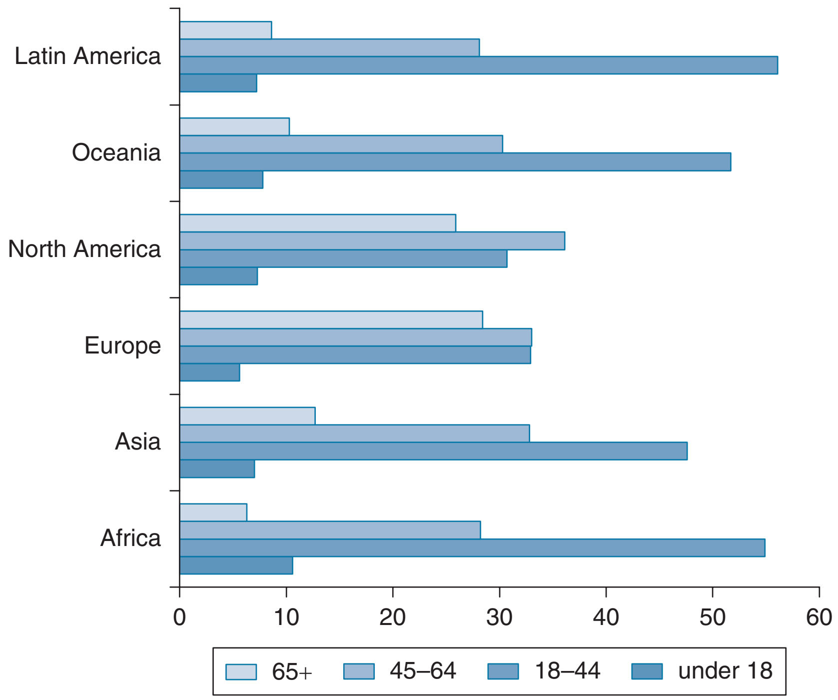
5. Support does vary by political party. The majority of strong Democrats (58.1%) and Independents (66%) agree/strongly agree with the statement. The group with the lowest percentage of agreement is Strong Republicans at 49%. The percentage disagreeing with the statement is highest among Strong Republicans (36.7%) compared with 12.3% of Strong Democrats and 11.3% of Independents.

7. The group with the largest increase in voting rates is blacks, from 53% in 1996 to 66.2% in 2012. Blacks are the only group that did not experience a decline in voting rates for the years presented. Hispanic voting rates exceeded the voting rates for Asians in 2000 and remained higher than Asians through 2012. Hispanics and Asians have the lowest voting rates for all groups. As noted in the exercise, in the 2012 presidential election, blacks had the highest voting rates for all groups, followed by non-Hispanic whites, Hispanics, and Asians. White voting rates declined by 2% from 2008 to 2012. The highest voting rate for whites was in 2004 (67.2%), in 2008 for Hispanics (49.9%) and for Asians (47.6%).

9. If we identify younger Americans as those in the 18 to 24 and 25 to 44 age-groups and older Americans in the 45–64, 65–74, and 75 and over categories—the data indicate that as age increases, so does the percentage of voting in a Presidential election. The group with the highest percentage of voting is the 65- to 74-year olds, with 73.5% voting. The percentage drops for the 75 and over age-group, but is still higher than the reported percentages for the age-groups: 18–24, 25–44, and 45–64.

11. Overall, the highest percentage of smokers are in the 12th-grade category; the lowest are students in the 8th grade. The highest percentage of daily smokers for all grades is between 1996 and 1997 with percentages declining through 2014. (There are no data for 8th and 10th graders pre-1990.) Since 2012, the percentage of students smoking daily was at 10% or below.

13.



a. For Africa, Asia, Oceania, and Latin America, the largest age-group is 18 to 44 years. For Europe and North America, the age composition is slightly older; individuals aged 45 to 64 years are the largest age-group for both.

b. We display the data in a vertical bar graph. We selected a bar graph because country of origin (the basis of the percentage calculation) is nominal.

CHAPTER 3

1.

a. Mode = Routine (f = 379)

b. Median = Routine

c. Based on the mode and median for this variable, most respondents indicate that their lives are "routine."

d. A mean score could not be interpreted for this variable. A mean would have no meaning for a nominal measurement.

3.

a. Interval ratio. The mode can be found in two ways: by looking either for the highest frequency (14) or the highest percentage (43.8%). The mode is the category that corresponds to the value "40 hours worked last week." The median can be found in two ways: by using either the frequencies column or the cumulative percentages.

Using Frequencies	Using Cumulative Percentages
$\frac{N+1}{2}=\frac{32+1}{2}=16.5\text{th case}$	Notice that 34.4% of the observations fall in or below the "32 hours worked last week" category; 78.1% fall in or below the "40 hours worked last week" category.
Starting with the frequency in the first category (1), add up the frequencies until you find where the 16th and 17th cases fall. Both these cases correspond to the category "40 hours worked last week," which is the median.	The 50% mark, or the median, is located somewhere within the "40 hours worked last week" category. So the median is "40 hours worked last week."

b. Since the median is merely a synonym for the 50th percentile, we already know that its value is "40 hours worked last week."

25th percentile = (32 × 0.25) = 8th case = 30 hours worked last week.

75th percentile = (32 × 0.75) = 24th case = 40 hours worked last week

5.

breakfastV8526 2014 T02 OFTN EAT BRKFST F2 * raceV1070 2014 RACE--B/W/H F1234 Crosstabulation

			raceV1070 2014 RACE--B/W/H F1234			
			1 BLACK:(1)	2 WHITE:(2)	3 HISPANIC:(3)	Total
breakfastV8526 2014 T02 OFTN EAT BRKFST F2	1 NEVER:(1)	Count	4	14	5	23
		% within raceV1070 2014 RACE--B/W/H F1234	9.3%	6.4%	7.0%	6.9%
	2 SELDOM:(2)	Count	11	22	9	42
		% within raceV1070 2014 RACE--B/W/H F1234	25.6%	10.1%	12.7%	12.7%
	3 SOMETIMES:(3)	Count	10	43	21	74
		% within raceV1070 2014 RACE--B/W/H F1234	23.3%	19.7%	29.6%	22.3%
	4 MOST DAYS:(4)	Count	2	25	9	36
		% within raceV1070 2014 RACE--B/W/H F1234	4.7%	11.5%	12.7%	10.8%
	5 NEARLY EVERY DAY:(5)	Count	7	30	2	39
		% within raceV1070 2014 RACE--B/W/H F1234	16.3%	13.8%	2.8%	11.7%
	6 EVERYDAY:(6)	Count	9	84	25	118
		% within raceV1070 2014 RACE--B/W/H F1234	20.9%	38.5%	35.2%	35.5%
Total		Count	43	218	71	332
		% within raceV1070 2014 RACE--B/W/H F1234	100.0%	100.0%	100.0%	100.0%

a.

	Mode	Median
Black	Seldom ($f = 11$)	Sometimes
White	Everyday ($f = 84$)	Nearly everyday
Hispanic	Everyday ($f = 25$)	Most days

b. Teens' breakfast habits vary by race/ethnicity. Out of the three racial/ethnic groups, black students were more likely to report seldom or sometimes eating breakfast. On the other hand, white and Hispanic students eat breakfast more frequently. The mode for white and Hispanic students is everyday.

7. We begin by multiplying each household size by its frequency.

Household Size	Frequency	Frequency × $Y (fY)$
1	381	381
2	526	1,052
3	227	681
4	200	800

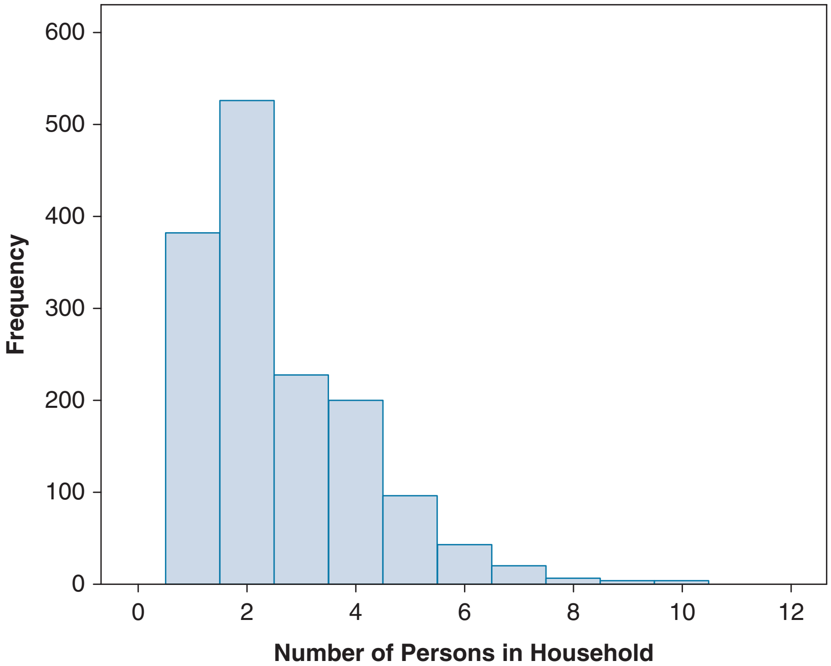
Household Size	Frequency	Frequency × Y(*f* Y)
5	96	480
6	42	252
7	19	133
8	5	40
9	2	18
10	2	20
Total	$N = 1{,}500$	$\Sigma fY = 3{,}857$

$$\bar{Y} = \frac{\Sigma f\,Y}{N} = \frac{3{,}857}{1{,}500} = 2.57$$

The mean number of people per household is 2.57.

9.

a. There appear to be a few outliers (i.e., extremely high values); this leads us to believe that the distribution is skewed in the positive direction.



b. The median can be found in two ways: by using either the frequencies column or the cumulative percentages. The data are in frequencies; we'll use those to solve the median. Because the median (2) is less than the mean (2.57), we can conclude that the distribution is skewed in a positive direction. Our answer to Question 9a is further supported.

Using Frequencies
$$\frac{N+1}{2}=\frac{1{,}500+1}{2}=750.5\text{th case}$$
Starting with the frequency in the first category (381), add up the frequencies until you find where the 750th and 751st cases fall. Both these cases correspond to the category "2," which is the median.

11. The mean and the median represent a precise statistical middle. The mean is often referred to as the "arithmetic middle," by definition, summing everyone's income and dividing the total by the total number of people. The mean is sensitive to extremes, very low or high values and so when we consider income, the preferred measure is the median. The median is the midpoint of all collected incomes, representing the exact point where 50% of all cases are either above or below. Because Clinton and Sanders' middle-class income amount is higher than the U.S. Census estimated mean or median, their definition of middle class is not based on the statistical middle. Are they operationalizing a middle-class life style, one that includes home and car ownership, occupational status, and wealth?

13.

a. The data are reordered to calculate the median.

Infant Mortality Rates
2.52
3.43
4.65
5.87
10.41
14.58
15.61
18.87
26.11
58.19
115.08
Median = 14.58, 6th case
Mean = 275.32/11 = 25.03

b. The mean is greater than median, indicating a positively skewed distribution.

CHAPTER 4

1.

a. The table reveals seven response categories for political views.

b. The sum of the squared percentages, ΣPct^2, is equal to 2,301.52.

Political Views	Percentage (%)	Percentage Squared ($\%^2$)
Extremely liberal	3.6	12.96
Liberal	12.7	161.29
Slightly liberal	11.1	123.21
Moderate	39.5	1560.25
Slightly conservative	14.4	207.36
Conservative	14.9	222.01
Extremely conservative	3.8	14.44
Total	100.0	$\sum = 2{,}301.52$

c. Using the formula, we calculate the IQV as follows:

$$\text{IQV} = \frac{K(100^2 - \sum Pct^2)}{100^2(K-1)} = \frac{7(100^2 - 2{,}301.52)}{100^2(7-1)} = \frac{53{,}889.36}{60{,}000} = 0.90$$

The calculated IQV is close to 1 and suggests that Americans are fairly diverse in their political views.

3.

a. The range of convictions in 1990 is (583 – 79) = 504. The range of convictions in 2009 is (426 – 102) = 324. The range of convictions is larger in 1990 than in 2009.

b. The mean number of convictions is 295.67 in 1990 and 261.67 in 2009.

c.

1990			
Govt. Level	No. of Convictions	$(Y - \bar{Y})$	$(Y - \bar{Y})^2$
Federal	583	287.33	82,558.53
State	79	–216.67	46,645.89
Local	225	–70.67	4,994.25

(Continued)

(Continued)

Govt. Level	No. of Convictions	$(Y - \bar{Y})$	$(Y - \bar{Y})^2$
Total	887		134,498.67
	$\bar{Y} = 295.67$		
$s = \sqrt{s^2} = \sqrt{\frac{\Sigma(Y-\bar{Y})^2}{N-1}} = \sqrt{\frac{134,498.67}{2}} = 259.32$			

2009			
Govt. Level	**No. of Convictions**	$(Y - \bar{Y})$	$(Y - \bar{Y})^2$
Federal	426	164.33	27,004.35
State	102	–159.67	25,494.50
Local	257	–4.67	21.80
Total	785		52,520.65
	$\bar{Y} = 261.67$		
$s = \sqrt{s^2} = \sqrt{\frac{\Sigma(Y-\bar{Y})^2}{N-1}} = \sqrt{\frac{52,520.65}{2}} = 162.05$			

d. The standard deviation is larger in 1990 than in 2009, thus indicating more variability in number of convictions in 1990 than in 2009. This supports our results from 3a.

5.

a. The range of projected increase in the elderly population for the Western states is 36.2%. The range of percent increase for the Midwestern states is 9.8%. The Western states have a much larger range.

b. The IQR for the Western states is 17.3%. The IQR for the Midwestern states is 3.7%. Again, the value for the Western states is greater.

c. There is great variability in the projected increase in the elderly population in Western states, chiefly caused by the large increases in Nevada, Arizona, Wyoming, and Alaska, as measured by either the range or the IQR.

7.

a. The range is 3.6 (6.5 – 2.9). The 25th percentile, 3.05, means that 25% of cases fall below 3.05 divorce rate per 1,000 population. Likewise, the 75th percentile means that 75% of all cases fall below 4.6 divorce rate per 1,000 population.

25th percentile	10(0.25) = 2.5th case	So (3.0 + 3.1)/2 = 3.05
75th percentile	10(0.75) = 7.5th case	So (4.5 + 4.7)/2 = 4.6

The IQR is thus 4.6 – 3.05 = 1.55. Both measures of variability are appropriate, but the range is somewhat better, as the value for the IQR is fairly small. In other words, the range gives us a better picture of the variability of divorce rates for all states in our sample.

b.

State	Divorce Rate per 1,000 Population	$Y - \bar{Y}$	$(Y - \bar{Y})^2$
Alaska	4.3	0.2	0.04
Florida	4.7	0.6	0.36
Idaho	4.9	0.8	0.64
Maine	4.5	0.4	0.16
Maryland	3.1	–1.0	1.00
Nevada	6.5	2.4	5.76
New Jersey	3.0	–1.1	1.21
Texas	3.3	–0.8	0.64
Vermont	3.8	–0.3	0.09
Wisconsin	2.9	–1.2	1.44
Total	41	0.00	11.34

$$\bar{Y} = \frac{\Sigma Y}{N} = \frac{41}{10} = 4.1$$

$$s = \sqrt{s^2} = \sqrt{\frac{\Sigma\left(Y - \bar{Y}\right)^2}{N-1}} = \sqrt{\frac{11.34}{9}} = 1.12$$

c. Divorce rates may vary by state due to factors such as variation in religiosity, state policy (i.e., no fault divorce laws), or employment opportunities.

9.

a. The mean number of crimes is 3,038.9 and the standard deviation is 583.004. The mean amount of dollars (in millions) spent on police protection is $1,703.95 and the standard deviation is $1,895.214.

b. Because the number of crimes and police protection expenditures are measured according to different scales, it isn't appropriate to directly compare the mean and standard deviation for one variable with the other. But we can talk about each distribution separately. We know from examining the mean (3,038.90) and standard deviation (583.00) for the number of crimes that the standard deviation is large, indicating a wide dispersion of scores from the mean. For the number of crimes, states such as Missouri and South Dakota contribute more to its variability because they have values far from the mean (both above and below). With respect to police protection expenditures, we can see that there is a large dispersion from the mean of $1,703.95, as the standard deviation is $1,895.21. States such as New York and North Dakota contribute more to its variability because they have values far from the mean (both above and below).

c. Among other considerations, we need to consider the economic conditions in each state. A downturn in the local and state economy may play a part in the number of crimes and police expenditures per capita.

11.

a. Type of paid work is a nominal variable. The appropriate measure of variability would be the index of qualitative variation (IQV).

b.

Grade 8		
Type of Work	Percentage (%)	Percentage Squared (%²)
Lawn work	28	784
Food service	3	9
Babysitting	37	1369
Other	32	1,024
Total	100.0%	$\Sigma = 3186$

$$\text{IQV} = \frac{K(100^2 - \sum Pct^2)}{100^2(K-1)} = \frac{4(100^2 - 3{,}186)}{100^2(4-1)} = \frac{27{,}256}{30{,}000} = 0.91$$

The IQV for 8th graders is 0.91.

Grade 10		
Type of Work	Percentage (%)	Percentage Squared (%²)
Lawn work	20	400
Food service	10	100
Babysitting	28	784
Other	42	1,764
Total	100.0%	$\Sigma = 3{,}048$

Type of Work	Percentage (%)	Percentage Squared (%²)
$$\text{IQV} = \frac{K(100^2 - \sum Pct^2)}{100^2(K-1)} = \frac{4(100^2 - 3{,}048)}{100^2(4-1)} = \frac{27{,}808}{30{,}000} = 0.93$$		

The IQV for 10th graders is 0.93.

c. Though both IQVs are more than 0.90, there is slightly more variation among 10th graders than 8th graders in the type of jobs they hold. The difference could be attributed to more employment options for older students. Younger students may be limited to the kind of work they can do (due to age, experience, and time), leading to more informal jobs, such as lawn work and babysitting.

13. Overall, Obama voters were younger, more educated, and attended religious services less than McCain voters. The youngest voters were female Obama voters at 50.99 years (s = 16.62), followed by male Obama voters, 51.71 years (s = 15.59). For education, males who voted for Obama had the highest mean of 14.84 (s = 3.07). Males who voted for McCain had 14.60 years of education (s = 2.41). McCain voters, both males and females, attended religious services more often than Obama voters. Mean scores were 3.93 for males (s = 2.80) and 4.64 for females (s = 2.76), indicating church attendance about once a month to 2 × 3 times per month. The standard deviations indicate a consistency in the distributions of education, age, and religious service attendance across all four groups. The largest standard deviations are for age, ranging from 15.61 to 16.62 years. These wide standard deviations indicate more dispersion around the mean age scores.

CHAPTER 5

1.

a. The Z score for a person who watches more than 8 hours per day:

$$Z = \frac{8 - 2.94}{2.60} = 1.95$$

b. We first need to calculate the Z score for a person who watches 5 hours per day:

$$Z = \frac{5 - 2.94}{2.60} = 0.79$$

The area between Z and the mean is 0.2852. We then need to add 0.50 to 0.2852 to find the proportion of people who watch television less than 5 hours per day. Thus, we conclude that the proportion of people who watch television less than 5 hours per day is 0.7852. This corresponds to 786 people (785.99 = 0.7852 × 1,001).

c. 5.54 television hours per day corresponds to a Z score of +1.

$$Y = \overline{Y} + Z(S_Y) = 2.94 + 1(2.60) = 5.54$$

d. The Z score for a person who watches 1 hour of television per day is –.75. The area between the mean and Z is 0.2734.

$$\frac{1-2.94}{2.60}=-.75$$

The Z score for a person who watches 6 hours of television per day is 1.18. The area between the mean and Z is .3810.

$$\frac{6-2.94}{2.60}=1.18$$

Therefore, the percentage of people who watch between 1 and 6 hours of television per day is 65.44% (0.2734 + 0.3810 = 0.6544 × 100).

3.

a. For an individual with 13.77 years of education, his or her Z score would be

$$Z=\frac{13.77-13.77}{3.07}=0.0$$

b. Since our friend's number of years of education completed is associated with the 60th percentile, we need to solve for Y. However, we must first use the logic of the normal distribution to find Z. For any normal distribution, 50% of all cases will fall above the mean. Since our friend is in the 60th percentile, we know that the area between the mean and our friend's score is 0.10. Similarly, the area beyond our friend's score is 0.40. We can now look in Appendix B column "B" for 0.10 or in column "C" for 0.40. We find that the Z associated with these values is 0.25. Now, we can solve for Y:

$$Y=\overline{Y}+Z(s)=13.77+0.25(307)=14.5375=14.54$$

c. Since we already know that the proportion between our number of years of education (13.77) and our friend's number of years of education (14.55) is 0.10, we can multiply N (1,500) by this proportion. Thus, 150 people have between 13.77 and 14.55 years of education.

5.

a. Among working-class respondents:

The Z score for a value of 12 is

$$Z=\frac{12-13.01}{2.91}=-0.35$$

The Z score for a value of 16 is

$$Z=\frac{16-13.01}{2.91}=1.03$$

You'll find the area between the Z scores and the mean under Column B. The total area between the scores is .1368 + .3485 = .4853. The proportion of working-class respondents with 12 to 16 years of education is .4853.

Among upper-class respondents:

The Z score for a value of 12 is

$$Z=\frac{12-15.44}{2.83}=-1.22$$

The Z score for a value of 16 is

$$Z = \frac{16 - 15.44}{2.83} = 0.20$$

The area between a Z of −1.22 and the mean is 0.3888. The area between a Z of 0.20 is 0.0793, so the total area between the scores is 0.3888 + 0.0793 = 0.4681. The proportion of upper-class respondents with 12 to 16 years of education is 0.4681.

A higher proportion of working-class respondents have 12 to 16 years of education than upper-class respondents.

b. For working-class respondents:

As previously calculated, the Z score for a value of 16 is 1.03. The area between a Z of 1.03 and the tail of the distribution (Column C) is 0.1515. So the probability of a working-class respondent having more than 16 years of education is 15.15%.

For middle-class respondents:

The Z score for a value of 16 is

$$Z = \frac{16 - 14.99}{2.93} = .35$$

The area between a Z of .35 and the tail of the distribution (Column C) is .3632. So the probability of a middle-class respondent having more than 16 years of education is 36.32%.

c. For lower-class respondents:

The Z score for a value of 10 is

$$Z = \frac{10 - 12.11}{2.83} = -0.75$$

The area beyond Z of 0.75 is .2266. So the probability of a lower-class respondent having less than 10 years of education is .2266 (or 22.66%).

d. If years of education is positively skewed, then the proportion of cases with high levels of education will be less than that for a normal distribution. This means, for example, that the probabilities associated with high levels of education will be smaller.

7.

a. The Z score of 150 is 3.33.

b. The area beyond 3.33 is 0.0004. The percentage of scores above 150 is 0.04%, a very small percentage.

c. The Z score for 85 is −1.0. The percentage of scores between 85 and 150 is 84.09% (0.3413 + 0.4996 = 0.8409).

d. Scoring in the 95th percentile means that 95% of the sample scored below this level. Identifying the 95th percentile can be calculated by this formula: 100 + 1.65(15) = 124.75. The IQ score that is associated with the 95th percentile is 124.75.

9.

a. About 0.1894 of the distribution falls above the Z score, so that is the proportion of crime incidents with more than two victims.

$$Z = \frac{2 - 1.28}{0.82} = 0.88$$

b. The area between the mean and the Z score is about 0.1331, so the total area above one victim is 0.50 + 0.1331 = 0.6331, or 63.31%.

$$Z = \frac{1 - 1.28}{0.82} = -0.34$$

c. The area between the mean and the Z score is about 0.4995, so the total area below four victims is 0.50 + 0.4995 = 0.9995.

$$Z = \frac{4 - 1.28}{0.82} = 3.32$$

11.

a. For a team with an APR score of 990

$$Z = \frac{990 - 981}{27.3} = 0.33$$

From Appendix B, the area beyond 0.33 is 0.3707 or about the 67th percentile. The team is not in the upper quartile.

b. The Z value which corresponds to a cutoff score with an area of about 0.25 toward the tail of the distribution is 0.67. This is translated into a cutoff score:

$$981 + 0.67(27.3) = 999.29.$$

c. The Z value is 0.67.

13. The 95th percentile corresponds to a Z score of 1.65.

Hungary

11.76 + 1.65 (2.91) = 16.56 years

Czech Republic

12.82 + 1.65 (2.29) = 16.60 years

Denmark

13.93 + 1.65 (5.83) = 23.55 years

France

14.12 + 1.65 (5.73) = 23.57 years

Ireland

15.15 + 1.65 (3.90) = 21.59 years

CHAPTER 6

1.

a. Although there are problems with the collection of data from all Americans, the census is assumed to be complete, so the mean age would be a parameter.

b. A statistic because it is estimated from a sample.

c. A statistic because it is estimated from a sample.

d. A parameter because the school has information on all employees.

e. A parameter because the school would have information on all its students.

3.

a. Assuming that the population is defined as all persons shopping at that shopping mall that day of the week, she is selecting a systematic random sample. A more precise definition might limit it to all persons passing by the department store at the mall that day.

b. This is a stratified sample because voters were first grouped by county, and unless the counties have the same number of voters, it is a disproportionate stratified sample because the same number is chosen from each county. We can assume that it was a probability sample, but we are not told exactly how the 50 voters were chosen from the lists. However, assuming that the population is defined as all Americans, this sort of sampling technique would qualify as nonprobability sampling.

c. This is neither a simple random sample nor a systematic random sample. It might be thought of as a sample stratified on last name, but even then, choosing the first 20 names is not a random selection process.

d. This is not a probability sample. Instead, it is a purposive sample chosen to represent a cross-section of the population in New York City.

5. The relationship between the standard error and the standard deviation is $\sigma_{\bar{Y}} = \sigma / \sqrt{N}$. Since σ is divided by $\sqrt{N}$, $\sigma_{\bar{Y}} = \sigma / \sqrt{N}$ must always be smaller than σ, except in the trivial case where $N = 1$. Theoretically, the dispersion of the mean must be less than the dispersion of the raw scores. This implies that the standard error of the mean is less than the standard deviation.

7.

a. These polls are definitely not probability samples. No sampling is done by the television station to choose who calls the 800 number.

b. The population is all those people who watch the television channel and see the 800 number advertised.

9.

a. This is not a random sample. The students eating lunch on Tuesday are not necessarily representative of all students at the school, and you have no way of calculating the probability of inclusion of any student. Many students might, for example, rarely eat lunch at the cafeteria and, therefore, have no chance of being represented in your sample. The fact that you selected *all* the students eating lunch on Tuesday makes your selection appear to be a census of a population, but that isn't true either unless all the students ate at the cafeteria on Tuesday.

b. This is a systematic random sample because names are drawn systematically from the list of all enrolled students.

c. This would seem to be a systematic random sample as in (b), but it suffers from the same type of defect as the cafeteria sample. Unless all students pass by the students union, using that location as a selection criterion means that some students have no chance of being selected (but you don't know which ones). Samples are often drawn this way in shopping malls by choosing a central location from which to draw the sample. It is reasonable to assume that a sufficiently representative mix of shoppers will pass by a central location during any one period.

d. The second procedure (selecting every 10th student from the list of all enrolled students) is the best option because it uses a random sampling method.

11.

a. Mean = 5.3 (53/10); standard deviation = 3.27.

b. Here are 10 means from random samples of size 3: 6.33, 5.67, 3.33, 5.00, 7.33, 2.33, 6.00, 6.33, 7.00, 3.00.

c. The mean of these 10 sample means is 5.23. The standard deviation is 1.76. The mean of the sample means is very close to the mean for the population. The standard deviation of the sample means is much less than the standard deviation for the population. The standard deviation of the means from the samples is an estimate of the standard error of the mean we would find from one random sample of size 3.

CHAPTER 7

1.

a. The estimate at the 90% confidence level is 22.82% to 23.18%. This means that there are 90 chances out of 100 that the confidence interval will contain the true population percentage of victims in the American population.

Due to the large sample size, we converted the proportions to percentages, subtracting from 100, rather than 1.

$$\text{Standard error} = \sqrt{\frac{(23)\,(100-23)}{160{,}040}} = 0.105 = 0.11$$

$$\begin{aligned}\text{Confidence interval} &= 23 \pm 1.65(0.11)\\ &= 23 \pm 0.18\\ &= 22.82 \text{ to } 23.18\end{aligned}$$

b. The true percentage of crime victims in the American population is somewhere between 22.72% and 23.28% based on the 99% confidence interval. There are 90 chances out of 100 that the confidence interval will contain the true population percentage of crime victims.

$$\begin{aligned}\text{Confidence interval} &= 23 \pm 2.58(0.11)\\ &= 23 \pm 0.28\\ &= 22.72 \text{ to } 23.28\end{aligned}$$

3.

a. For Canadians

$$s_p = \sqrt{\frac{(0.51)(1-0.51)}{1{,}004}} = 0.02$$

$$\text{Confidence interval} = 0.51 \pm 1.96(0.02)$$
$$= 0.51 \pm 0.04$$
$$= 0.47 \text{ to } .55$$

b. For Americans

$$\text{Confidence interval} = 0.45 \pm 1.96(0.02)$$
$$= 0.45 \pm 0.04$$
$$= 0.39 \text{ to } 0.49$$

c. Based on the calculated 95% confidence interval, the majority of Americans do not believe climate change is a serious problem. The true percentage of Americans who believe climate change is a serious problem is under 50%, somewhere between 39% and 49%, based on this Pew Research Center sample. On the other hand, it is possible that the majority of Canadians believe climate change is a serious problem. We can be 95% confident that the true percentage of Canadians is somewhere between 47% and 55%.

5.

Due to the large sample size, we converted the proportion to full percentages, subtracting from 100 (rather than 1).

$$\text{Confidence interval} = 51 \pm 1.96(0.67)$$
$$= 49.69\% \text{ to } 52.31\%$$

We set the interval at the 95% confidence level. However, no matter whether the 90%, 95%, or 99% confidence level is chosen, the calculated interval includes values below 50% for the vote for a Republican candidate. Therefore, you should tell your supervisors that it would not be possible to declare a Republican candidate the likely winner of the votes coming from men if there was an election today because it seems quite possible that less than a majority of male voters would support her or him.

7.

a.

$$s_p = \sqrt{\frac{(0.64)(1-0.64)}{1,403}} = 0.01$$

$$\text{Confidence interval} = 0.64 \pm 1.96(0.01)$$
$$= 0.64 \pm 0.02$$
$$= 0.62 \text{ to } 0.66$$

b. Based on our answer in 7a, we know that a 90% confidence interval will be more precise than a 95% confidence interval that has a lower bound of 62% and an upper bound of 66%. Accordingly, a 90% confidence interval will have a lower bound that is greater than

62% and an upper bound that is less than 66%. Additionally, we know that a 99% confidence interval will be less precise than what we calculated in 7a. Thus, the lower bound for a 99% confidence interval will be less than 62% and the upper bound will be greater than 66%.

9.

Country	Mean	Standard Error	Confidence Interval
France	14.12	$5.73/\sqrt{975} = 0.18$	$14.12 + 0.18(1.65) = 14.42$ $14.12 - 0.18(1.65) = 13.82$
Japan	12.48	$2.53/\sqrt{528} = 0.11$	$12.48 + 0.11(1.65) = 12.66$ $12.48 - 0.11(1.65) = 12.30$
Croatia	12.18	$2.71/\sqrt{480} = 0.12$	$12.18 + 0.12(1.65) = 12.38$ $12.18 - 0.12(1.65) = 11.98$
Turkey	9.15	$11.98/\sqrt{783} = 0.43$	$9.15 + 0.43(1.65) = 9.86$ $9.15 - 0.43(1.65) = 8.44$

11.

For Republicans

$$s_p = \sqrt{\frac{(0.18)(1-0.18)}{446}} = 0.02$$

$$\text{Confidence interval} = 0.18 \pm 1.96(0.02)$$

$$= 0.18 \pm 0.04$$

$$= 0.14 \text{ to } 0.22$$

For Democrats

$$s_p = \sqrt{\frac{(0.15)(1-0.15)}{522}} = 0.02$$

$$\text{Confidence interval} = .15 \pm 1.96(0.02)$$

$$= 0.15 \pm 0.04$$

$$= 0.11 \text{ to } 0.19$$

13.

a. For those who thought that homosexual relations were always wrong:

$$s_p = \sqrt{\frac{(0.40)(1-.40)}{950}} = 0.02$$

$$\text{Confidence interval} = 0.40 \pm 1.96(.02)$$

$$= 0.40 \pm 0.04$$

$$= 0.36 \text{ to } 0.44$$

For those who thought that homosexual relations were not wrong at all:

$$s_p = \sqrt{\frac{(0.49)(1-0.49)}{950}} = 0.02$$

$$\text{Confidence interval} = 0.49 \pm 1.96(.02)$$
$$= 0.49 \pm 0.04$$
$$= 0.45 \text{ to } 0.53$$

b.

$$s_p = \sqrt{\frac{(0.10)(1-0.10)}{950}} = 0.01$$

$$\text{Confidence interval} = 0.10 \pm 1.96(.01)$$
$$= 0.10 \pm 0.02$$
$$= 0.08 \text{ to } 0.12$$

CHAPTER 8

1.

a. H_0: $\mu = 13.5$ years; H_1: $\mu < 13.5$ years.

b. The Z value obtained is -4.19. The p value for a Z of -4.19 is less than .001 for a one-tailed test. This is less than the alpha of .01, so we reject the null hypothesis and conclude that the doctors at the HMO do have less experience than the population of doctors at all HMOs.

3.

a. Two-tailed test, $\mu \neq \$53{,}657$; null hypothesis, $\mu = \$53{,}657$

b. One-tailed test, $\mu > 3.2$; null hypothesis, $\mu = 3.2$

c. One-tailed test, $\mu_1 < \mu_2$; null hypothesis, $\mu_1 = \mu_2$

d. Two-tailed test, $\mu_1 \neq \mu_2$; null hypothesis, $\mu_1 = \mu_2$

e. One-tailed test, $\mu_1 > \mu_2$; null hypothesis, $\mu_1 = \mu_2$

f. One-tailed test, $\mu_1 < \mu_2$; null hypothesis, $\mu_1 = \mu_2$

5.

a. Research hypothesis, $\mu \neq 37.2$; null hypothesis, $\mu = 37.2$

b. The t obtained is -29.36 and its p level is <.001 (it is greater than the last reported t critical of 3.291).

$$t = \frac{37.2 - 50.12}{17.07 / \sqrt{1490}} = \frac{-12.92}{.44} = -29.36$$

c. We conclude that we can reject the null hypothesis in favor of the research hypothesis. There is a difference between the mean age of the GSS sample and the mean age of all

American adults. Relative to age, the GSS sample is not representative of all American adults (the GSS sample is significantly older).

7.

a. The appropriate test statistic is t test for sample means.

b.

$$t = \frac{2.35 - 3.05}{.18} = -3.89$$

$$\text{Standard error} = \sqrt{\frac{188(1.21)^2 + 60(1.05)^2}{(189+61)-2}}\sqrt{\frac{189+61}{(189)(61)}} = (1.17)(.15) = 0.18$$

The t obtained of −3.89 is greater than the t critical of −1.645. We reject the null hypothesis of no difference. College graduates are more likely to indicate that being Christian is “not very important,” whereas high school graduates indicate that being Christian is “fairly important.”

c. For a two-tailed test, the t critical would be 1.96. The t obtained is still greater. We would reject the null hypothesis of no difference.

9.

a. “Less than” indicates a one-tailed test.

b. $Z = 3.00$. with a significance of .001. We can reject the null hypothesis and conclude that the proportion of males who believe in the historical importance of the election of a woman for president is significantly less than the proportion of female voters who believe the same.

$$\frac{.55 - .65}{.02} = -5.0$$

c. The significance of −5.00 is less than .01 (.0014 < .001). The decision to reject the null hypothesis does not change.

11. Older individuals, aged 50 to 59 years, gave more money in the past year than younger adults aged 30 to 39 years. However, the difference in giving is not significant. The t obtained is −.800 (equal variances assumed) with a probability of .425 (>.05 alpha).

13.

a. Yes, there is a significant difference between the average number of relaxation hours for married men and women. Married women have significantly less relaxation hours per day than men in the GSS 2014 sample, a difference of .68 hours (3.56 – 2.88). The t obtained of 2.225 is significant at the .025 level (less than our alpha of .05).

b. If alpha was changed to .01, we would fail to reject the null hypothesis of no difference. The probability of the t obtained is .025 > .01.

CHAPTER 9

1.

a. The independent variable is sex; the dependent variable is fear of walking alone at night.

Fear of Walking Alone at Night	Sex	
	Male F (%)	Female F (%)
Yes	2 (22%)	8 (73%)
No	7 (78%)	3 (27%)

b. Approximately 73% of women are afraid to walk alone in their neighborhoods at night, whereas approximately 22% of men said the same. This amounts to about a 51% difference between women and men who are not afraid to walk alone at night, indicating a strong relationship. It is important to keep in mind that our small sample size limits the generalizability of these results.

c. There is a relationship between homeownership and fear of walking alone at night. The majority of homeowners (56%) were not afraid of walking alone at night in their neighborhood. Among those who were not homeowners, the majority (55%) were afraid of walking alone at night.

Fear of Walking Alone at Night	Home Ownership	
	Yes F (%)	No F (%)
Yes	4 (44%)	6 (55%)
No	5 (56%)	5 (45%)

3.

a. Based on the student's argument the independent variable is *attitude toward homosexual relations* and the dependent variable is *political views*.

b. 285/645 = 44%

c. Those who believe that homosexuality is always wrong are more likely to be conservative (51%) than moderate (34%) or liberal (15%). On the other hand, those who believe homosexuality is not wrong at all are more likely to report liberal (41%) or moderate (40%) views than conservative (18%) ones.

5. The relationship is weak between race and the frequency of being drunk in the past 12 months. The majority of students are likely to report not being drunk in the past 12 months, at least 86% of each racial group. The percentage of students being drunk at least 3 or more times is highest for whites (7%), followed by Hispanic (6%), and black (2%) students.

Drunk in the Last 12 Months	Race			Total
	Black	White	Hispanic	
None	75	282	119	476
	90%	86%	90%	

(Continued)

(Continued)

Drunk in the Last 12 Months	Race			Total
	Black	White	Hispanic	
1–2 times	6	23	5	34
	7%	7%	4%	
3–5 times	0	13	3	16
		4%	2%	
6 or more times	2	11	5	18
	2%	3%	4%	
Total	83	329	132	544

7. Female seniors have higher educational expectations than male seniors. For example, 73.9% (32.6 + 41.3) of female students expected to complete a bachelor's degree or higher. This is higher than the combined percentage for male students: 63.3% (34.4 + 28.9).

9. Yes, there is a relationship between political party affiliation and attitudes toward the Affordable Care Act. The majority of physicians who reported being Republican or Other party were strongly against or against the Affordable Care Act. The largest reporting percentage was among Republicans (84.07% = 31.08 + 52.99). Only 47.2% (11.20 + 36) of Democrats were strongly against or against the act.

11. The data indicate a positive relationship between students' educational expectations and parental education. The percentage of students indicating a bachelor's degree or higher increases as the parents' educational level increases: from 52% of students with parents with a high school degree or less to 86.1% of students with parents who completed a graduate/ professional degree.

13. In contrast with male students, female students are more likely to report not being drunk at all (at least 82% of each racial group). According to the data, Hispanic females (9%) are more likely to report being drunk three or more times in the last 12 months than white (7%) or black (6%) females.

Drunk in the Last 12 Months	Race			Total
	Black	White	Hispanic	
None	76	286	100	462
	87%	83%	82%	
1–2 times	6	33	11	50
	7%	10%	9%	
3–5 times	4	12	7	23
	5%	3%	6%	

Drunk in the Last 12 Months	Race			Total
	Black	White	Hispanic	
6 or more times	1	14	4	19
	1%	4%	3%	
Total	87	345	122	554

CHAPTER 10

1.

a. Degrees of freedom = (2 − 1)(2 − 1) = 1

b. Chi-square = 29.01 (with Yates's correction is 28.18). The probability of our obtained chi-square is less than our alpha (and less than 0.001). We can reject the null hypothesis and conclude that gender and fear of walking alone at night are dependent. A higher percentage of women (40%) than men (22%) report being afraid.

Sex and FEAR	f_o	f_e	$f_o - f_e$	$(f_o - f_e)^2$	$\frac{(f_o - f_e)^2}{f_e}$
Men/yes	77	111.96	–34.96	1222.20	10.92
Men/no	270	235.04	34.96	1222.20	5.20
Women/yes	175	140.04	34.96	1222.20	8.73
Women/no	259	293.96	–34.96	1222.20	4.16
$\chi^2 = 29.01$					

With the Yates correction:

Sex and FEAR	$\lvert f_o - f_e \rvert$	$(\lvert f_o - f_e \rvert - .50)^2$	f_e	$\frac{(\lvert f_o - f_e \rvert - .5)^2}{f_e}$
Men/yes	34.96	(34.46)2 = 1187.49	111.96	10.61
Men/no	34.96	(34.46)2 = 1187.49	235.04	5.05
Women/yes	34.96	(34.46)2 = 1187.49	140.04	8.48
Women/no	34.96	(34.46)2 = 1187.49	293.96	4.04
Total				28.18

c. If α were changed to .01, we would still reject the null hypothesis. The probability of our obtained chi-square is still less than alpha.

d. The lambda is 0. There is no proportional reduction of error using sex to predict fear of walking alone at night.

3.

a. A slightly higher percentage of blacks, 34.2% (38/111), report being afraid to walk alone at night. Among whites, the percentage is 31.6% (190/601).

b. Regardless of race, women are more likely than men to report being afraid to walk alone at night. The percentage of white women indicating that they are afraid is slightly higher than black women, 40.4% (129/319) versus 39% (30/77).

c.

Whites, $\chi^2 = 23.62$; we reject the null hypothesis.

Blacks, $\chi^2 = 1.85$; we fail to reject the null hypothesis.

For Whites

Sex and FEAR	f_o	f_e	$f_o - f_e$	$(f_o - f_e)^2$	$\frac{(f_o - f_e)^2}{f_e}$
Men/yes	61	89.15	–28.15	792.42	8.89
Men/no	221	192.85	28.15	792.42	4.12
Women/yes	129	100.85	28.15	792.42	7.86
Women/no	190	218.15	–28.15	792.42	3.63
$\chi^2 = 24.50$					

With the Yates's correction, the chi-square is 23.62, as it is shown below:

Sex and FEAR	$\lvert f_o - f_e \rvert$	$(\lvert f_o - f_e \rvert - .50)^2$	f_e	$\frac{(\lvert f_o - f_e \rvert - .5)^2}{f_e}$
Men/yes	28.15	764.52	89.15	8.58
Men/no	28.15	764.52	192.85	3.96
Women/yes	28.15	764.52	100.85	7.58
Women/no	28.15	764.52	218.15	3.50

For Blacks

Sex and FEAR	f_o	f_e	$f_o - f_e$	$(f_o - f_e)^2$	$\frac{(f_o - f_e)^2}{f_e}$
Men/yes	8	11.64	–3.64	13.25	1.14
Men/no	26	22.36	3.64	13.25	.59

Sex and FEAR	f_o	f_e	$f_o - f_e$	$(f_o - f_e)^2$	$\frac{(f_o - f_e)^2}{f_e}$
Women/yes	30	26.36	3.64	13.25	.50
Women/no	47	50.64	–3.64	13.25	.26
$\chi^2 = 2.49$					

With the Yates's correction, the chi-square is 1.85:

Sex and FEAR	$\lvert f_o - f_e \rvert$	$(\lvert f_o - f_e \rvert - .50)^2$	f_e	$\frac{(\lvert f_o - f_e \rvert - .5)^2}{f_e}$
Men/yes	3.64	9.86	11.64	0.85
Men/no	3.64	9.86	22.36	0.44
Women/yes	3.64	9.86	26.36	0.37
Women/no	3.64	9.86	50.64	0.19

5.

a. We will make 2,973 errors, because we predict that all victims fall in the modal category (white). $E_1 = 6{,}084 - 3{,}111 = 2{,}973$.

b. For white offenders, we could make 373 errors; for black offenders, 493 errors; and for other offenders, we would make 42 errors. $E_2 = 908$.

c. The proportional reduction in error is then (2,973 – 908)/2,973 = .6946. This indicates a very strong relationship between the two variables. We can reduce the error in predicting victim's race based on race of offender by 69.46%.

7.

Race/First-Generation College Status	f_o	f_e	$f_o - f_e$	$(f_o - f_e)^2$	$\frac{(f_o - f_e)^2}{f_e}$
White/first	1,742	1,749.6	–7.6	57.76	0.03
White/nonfirst	2,392	2,384.4	7.6	57.76	0.02
Black/first	102	93.5	8.5	72.25	0.77
Black/nonfirst	119	127.5	–8.5	72.25	0.57
Native American/first	41	36.4	4.6	21.16	0.58
Native American/nonfirst	45	49.6	–4.6	21.16	0.43
Hispanic/first	19	18.6	0.4	0.16	0.01
Hispanic/nonfirst	25	25.4	–0.4	0.16	0.01

(Continued)

(Continued)

Race/First-Generation College Status	f_o	f_e	$f_o - f_e$	$(f_o - f_e)^2$	$\frac{(f_o - f_e)^2}{f_e}$
Asian American/first	6	11.9	–5.9	34.81	2.93
Asian American/nonfirst	22	16.1	5.9	34.81	2.16
$\chi^2 = 7.51$					

Chi-square = 7.51, with 4 degrees of freedom [(2 – 1)(5 – 1) = 4].

We would fail to reject the null hypothesis. The probability of our obtained chi-square lies somewhere between 0.20 and 0.10, above our alpha level.

9. We would reject the null hypothesis. The chi-square obtained of 52.047 is significant at the .032 level (< .05 alpha). There is a relationship between degree and church attendance for these French respondents. Overall, as educational attainment increases, church attendance decreases.

11. The lambda of .051 for PRES12 and HLTHALL indicates a weak relationship. Only 5.1% of the error in predicting HLTHALL responses based on PRES12. Notice from the SPSS output that when PRES12 is defined as the dependent variable (HLTHALL is the independent variable), the lambda increases to .467.

    The gamma of –.198 indicates a weak negative relationship between CLASS and HLTHALL. If we rely on CLASS as an independent variable to predict HLTHALL, we would reduce our errors by 19.8%.

13. Gender: The model is significant at the .01 level, indicating a significant relationship between the variables. Though males contribute to more violent onset, in proportional terms, females exhibit a higher prevalence rate—18.32% of females exhibit violent onset compared with 11.71% of males.

    Age at first offense: The model is significant at the .01 level, indicating a significant relationship between age at first offense and violent onset. Violent onset is more likely among the group 14 years and older (14.74%) than those less than 14 years of age at first onset (9.67%).

CHAPTER 11

1.

$\bar{Y}_1 = 2.875$	$\bar{Y}_2 = 2.250$	$\bar{Y}_3 = 2.00$	$\bar{Y}_4 = 1.375$
$\sum Y_1 = 23$	$\sum Y_2 = 18$	$\sum Y_3 = 16$	$\sum Y_4 = 11$
$\sum Y_1^2 = 71$	$\sum Y_2^2 = 44$	$\sum Y_3^2 = 38$	$\sum Y_4^2 = 17$
$n_1 = 8$	$n_2 = 8$	$n_3 = 8$	$n_4 = 8$
$\bar{Y} = 2.125$			
$N = 32$			

$$SSB = 8(2.875 - 2.125)^2 + 8(2.250 - 2.125)^2 + 8(2.00 - 2.125)^2 + 8(1.375 - 2.125)^2$$
$$= 8(0.5625) + 8(.015625) + 8(.015625) + 8(.5625)$$
$$= 4.5 + .125 + .125 + 4.5$$
$$SSB = 9.25$$

$$df_b = 4 - 1$$
$$df_b = 3$$

$$\text{Mean square between} = 9.25/3 = 3.08$$

$$SSW = (71 + 44 + 38 + 17) - \left[(23^2/8) + (18^2/8) + (16^2/8) + (11^2/8)\right]$$
$$= 170 - (66.125 + 40.5 + 32 + 15.125)$$
$$= 170 - 153.75$$
$$SSW = 16.25$$

$$df_w = 32 - 4$$
$$= 28$$

$$\text{Mean square within} = 16.25/28 = 0.58$$

$$F = 3.08/0.58$$
$$= 5.31$$

Decision: If we set alpha at .05, F critical would be 2.95 ($df_1 = 3$ and $df_2 = 28$). Based on our F obtained of 5.31, we would reject the null hypothesis and conclude that at least one of the means is significantly different than the others. Upper-class respondents rate their health the highest (1.375), followed by middle- and working-class respondents (2.00 and 2.25, respectively) and lower-class respondents (2.875) on a scale where 1 = *excellent*, 4 = *poor*.

3.

$\bar{Y}_1 = 1.6$	$\bar{Y}_2 = 1.4$	$\bar{Y}_3 = 0.6$
$\sum Y_1 = 16$	$\sum Y_2 = 14$	$\sum Y_3 = 6$
$\sum Y_1^2 = 30$	$\sum Y_2^2 = 24$	$\sum Y_3^2 = 8$
$n_1 = 10$	$n_2 = 10$	$n_3 = 10$
$\bar{Y} = 1.2$		
$N = 30$		

$$SSB = 10(1.6 - 1.2)^2 + 10(1.4 - 1.2)^2 + 10(0.6 - 1.2)^2$$
$$= 10(0.16) + 10(0.04) + 10(0.36)$$
$$= 1.6 + 0.4 + 3.6$$
$$= 5.6$$

$$df_{\text{b}} = 3 - 1$$
$$= 2$$

$$\text{Mean square between} = 5.6/2 = 2.8$$

$$\begin{aligned} SSW &= (30 + 24 + 8) - (16^2/10) + (14^2/10) + (6^2/10) \\ &= 62 - (25.6 + 19.6 + 3.6) \\ &= 62 - 48.8 \\ &= 13.2 \end{aligned}$$

$$df_{\text{w}} = 30 - 3$$
$$df_{\text{w}} = 27$$

$$\text{Mean square within} = 13.2/27 = 0.488889$$

$$F = 2.8/0.49$$

$$= 5.71$$

Decision: If we set alpha at .01, F critical would be 5.49 (df_1 = 2 and df_2 = 27). Based on our F obtained of 5.71, we would reject the null hypothesis and conclude that at least one of the means is significantly different than the others. Respondents with no degree rate their church attendance highest (1.6), followed by respondents with a secondary degree (1.4) and then respondents with a university degree (0.6).

5.

$\bar{Y}_1 = 0.8$	$\bar{Y}_2 = 1.75$	$\bar{Y}_3 = 3.20$
$Y_1 = 4$	$Y_2 = 7$	$Y_3 = 16$
$\sum Y_1^2 = 6$	$\sum Y_2^2 = 15$	$\sum Y_3^2 = 54$
$n_1 = 5$	$n_2 = 4$	$n_3 = 5$
$\bar{Y} = 1.93$		
$N = 14$		

$$\begin{aligned} SSB &= 5(.8 - 1.93)^2 + 4(1.75 - 1.93)^2 + 5(3.20 - 1.93)^2 \\ &= 5(1.2769) + 4(.0324) + 5(1.6129) \\ &= 6.3845 + 0.1296 + 8.0645 \\ SSB &= 14.58 \end{aligned}$$

$$df_b = 3 - 1$$
$$df_b = 2$$

$$\text{Mean square between} = 14.58/2 = 7.29$$

$$
\begin{aligned}
SSW &= (6+15+54) - \left[\left(4^2/5\right) + \left(7^2/4\right) + \left(16^2/5\right)\right] \\
&= 75 - (3.2 + 12.25 + 51.2) \\
&= 75 - 66.65 \\
SSW &= 8.35
\end{aligned}
$$

$$
\begin{aligned}
df_w &= 14 - 3 \\
df_w &= 11
\end{aligned}
$$

$$\text{Mean square within} = 8.35/11 = 0.76$$

$$
\begin{aligned}
F &= 7.29/0.76 \\
F &= 9.59
\end{aligned}
$$

Decision. If we set alpha at .05, F critical would be 3.98 ($df_1 = 2$ and $df_2 = 11$). Based on our F obtained of 9.59, we would reject the null hypothesis and conclude that at least one of the means is significantly different from the others. The average number of moving violations is the highest for large-city respondents (3.2); medium-sized city residents are next (1.75), followed last by small-town respondents (0.8).

7. For each sociocultural resource, we would reject the null hypothesis. For social support, the obtained F ratio is 12.17, $p < .001$. Whites report the highest level of social support (2.85) while Non-Cuban Hispanics have the lowest (2.58). For religious attendance, the obtained F ratio is 56.43, $p < .001$. Church attendance is highest for African Americans and Non-Cuban Hispanics in the sample (3.94 and 3.37 on the 5-point scale).

9. Based on alpha = .01, we reject the null hypothesis of no difference. The average donation amount does vary by educational degree. The group with the highest average donation amount is graduate degree ($5590.61) followed by bachelor degree ($3397.40). The group with the lowest donation amount was less than high school graduates ($593.85).

11.

a. Yes, agreement to the statement does vary by how satisfied the individual is with his or her financial situation. The ANOVA model is significant at the .003 level (< .01 alpha). All group means are between agree (2) or neither (3), but the group most likely to agree with the statement is the group which is not at all satisfied with their financial situation. This group's mean score is 2.72, between agree and neither. For the satisfied and more or less satisfied with their financial situation, average scores are slightly above 3—neither agree or disagree.

b. Eta-squared is 14.662/501.637 = .029 = .03. Only 3% of the variation in IMMJOBS can be explained by satisfaction with finances.

13.

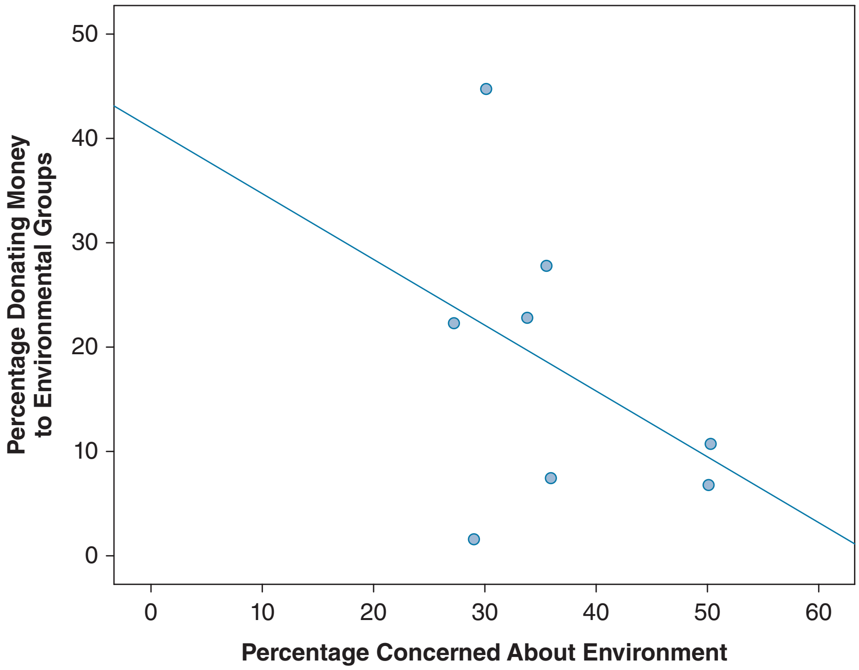
a. $df_b = k - 1 = 5 - 1 = 4$; $df_w = N - k = 254 - 5 = 249$

b. We would reject the null hypothesis for the three models. Students' perception of mentoring does vary by racial/ethnic identity. The most significant model is for the statement, "There are peer mentors who can advise me." Native American students have the highest level of agreement, followed by African American students. The lowest average score is for Asian students. The model for "I mentor other students" is significant at the .006 level. Native American students have the highest level of agreement, followed by African American students. The lowest average score is for Asian students. Finally, the model for "There are persons of color in administrative roles from whom I would seek mentoring at this institution" is significant at the .008 level. Native American students have the highest average level of agreement, followed by multiethnic students. The lowest score was reported by Hispanic students.

CHAPTER 12

1.

a. On the scatterplot below, the regression line has been plotted to make it easier to see the relationship between the two variables.



b. The scatterplot shows that there is a general linear relationship between the two variables. There is not a lot of scatter about the straight line describing the relationship. As the percentage of respondents concerned about the environment increases, the percentage of respondents donating money to environmental groups decreases.

c. The Pearson correlation coefficient between the two variables is –0.40. This is consistent with the scatterplot that indicated a negative relationship between being concerned about the environment and actually donating money to environmental groups.

Country	(1) Percentage Concerned X	(2) Percentage Donating Y	(3) $(X-\bar{X})$	(4) $(X-\bar{X})^2$	(5) $(Y-\bar{Y})$	(6) $(Y-\bar{Y})^2$	(7) $(X-\bar{X})(Y-\bar{Y})$
United States	33.8	22.8	−2.69	7.24	4.77	22.75	−12.83
Austria	35.5	27.8	−0.99	0.98	9.77	95.45	−9.67
The Netherlands	30.1	44.8	−6.39	40.83	26.77	716.63	−171.06
Slovenia	50.3	10.7	13.81	190.72	−7.33	53.73	−101.23
Russia	29.0	1.6	−7.49	56.10	−16.43	269.94	123.06
Philippines	50.1	6.8	13.61	185.23	−11.23	126.11	−152.84
Spain	35.9	7.4	−0.59	0.35	−10.63	113.00	6.27
Denmark	27.2	22.3	−9.29	86.30	4.27	18.23	−39.67
	$\sum X = 291.9$	$\sum Y = 144.2$	−0.02[a]	567.75	0.04[a]	1,415.84	−357.97

$$\text{Mean } X = \bar{X} = \frac{\sum X}{N} = \frac{291.9}{8} = 36.49$$

$$\text{Mean } Y = \bar{Y} = \frac{\sum Y}{N} = \frac{144.2}{8} = 18.03$$

$$\text{Variance } (Y) = s_Y^2 = \frac{\sum (Y-\bar{Y})^2}{N-1} = \frac{1{,}415.84}{7} = 202.26$$

$$\text{Standard deviation } (Y) = s_Y = \sqrt{202.26} = 14.22$$

$$\text{Variance } (X) = s_X^2 = \frac{\sum (X-\bar{X})^2}{N-1} = \frac{567.75}{7} = 81.11$$

$$\text{Standard deviation } (X) = s_X = \sqrt{81.11} = 9.01$$

$$\text{Covariance } (X, Y) = s_{XY} = \frac{\sum (X-\bar{X})(Y-\bar{Y})}{N-1} = \frac{-357.97}{7} = -51.14$$

$$r = \frac{s_{XY}}{s_X s_Y} = \frac{-51.14}{(9.01)(14.22)} = -0.40^a$$

Note: Answers may differ slightly due to rounding.

3.

a. The correlation coefficient is –0.45.

Country	(1) GNP per Capita	(2) Percentage Willing to Pay	(3)	(4)	(5)	(6)	(7)
State	X	Y	$(X-\bar{X})$	$(X-\bar{X})^2$	$(Y-\bar{Y})$	$(Y-\bar{Y})^2$	$(X-\bar{X})(Y-\bar{Y})$
United States	29.24	44.9	2.72	7.40	−1.64	2.69	−4.46
Ireland	18.71	53.3	−7.81	61.00	6.76	45.70	−52.80
The Netherlands	24.78	61.2	1.74	3.03	14.66	214.92	−25.51
Norway	34.31	40.7	7.79	60.68	−5.84	34.11	−45.49
Sweden	25.58	32.6	−0.94	0.88	−13.94	194.32	13.10
	$\sum X = 132.62$	$\sum Y = 232.7$	−0.02[a]	132.99	0.04[a]	491.74	−115.16

$$\text{Mean } Y = \bar{Y} = \frac{\sum Y}{N} = \frac{232.7}{5} = 46.54$$

$$\text{Variance } (X) = s_X^2 = \frac{\sum (X-\bar{X})^2}{N-1} = \frac{132.99}{4} = 33.25$$

$$\text{Standard deviation } (X) = s_X = \sqrt{33.25} = 5.77$$

$$\text{Variance } (Y) = s_Y^2 = \frac{\sum (Y-\bar{Y})^2}{N-1} = \frac{491.74}{4} = 122.94$$

$$r = \frac{s_{XY}}{s_X s_Y} = \frac{-28.79}{(5.77)(11.09)} = -0.45^a$$

Notes:

a. Answers may differ slightly due to rounding.

b. A correlation coefficient of −0.45 means that relatively high values of GNP are moderately negatively assoicated with low values of percentage of residents willing to pay higher prices to protect the environment.

5. The analysis reveals a negative relationship between years of education and number of children. The bivariate regression equation is $Y = 3.537 + -0.118X$. For each year increase in education, the number of children is predicted to decrease by 0.118. The model explains just 5% of the variance; however, based on the ANOVA F obtained, we can reject the null hypothesis that $r^2 = 0$.

7.

   a. The regression analysis confirms a positive relationship between years of education and total donations given in the past year. The F obtained is 10.578 (significant at .001). We can conclude that the relationship between the two variables is significant.

b. For respondent with 14 years of education: $2043.86

For respondent with 20 years of education: $3868.50

9.

a. For males: $Y = 9.768 + 0.355X$

For females: $Y = 9.770 + 0.367X$

b. For males, mother with 20 years of education: $9.768 + 0.355(20) = 16.87$

For females, mother with 20 years of education: $9.770 + 0.367(20) = 17.11$

c. The model for females has a slightly higher r^2. Mother's education explains 22% of the variance in female respondent education compared with the 20% explained for male respondent education. Based on the F-obtained statistic, both models are significant.

11.

a. Both hypotheses are confirmed.

The slope for education is 0.598. Holding age constant, for each year increase in education, Internet hours per week increases by 0.598.

The slope for age is –0.236. Holding years of education constant, for each year increase in age, Internet hours per week decreases by 0.236.

b. $Y = 14.395 + 0.598(X_1) + -0.236(X_2)$

$Y = 14.395 + 0.598(16) + -0.236(55) = 10.98$ Internet hours per week

c. $Y = 14.395 + 0.263(X_1) + -0.047(X_2)$

Education has the strongest effect on Internet hours per week (beta = .263).

d. The R^2 is 0.065. Education and age explain 6.5% of the variance in predicting Internet hours per week. This is a weak prediction model.

e. The correlation between Internet hours per week and age of respondent is –0.231, indicating a weak negative relationship. The correlation between Internet hours per week and education is 0.088, indicating a weak positive relationship. Finally, the correlation between age and education is –0.009, a weak negative relationship. The only significant correlation is the one between Internet hours and age.

13.

a. $Y = 3.91 + -0.115(X_1) + -0.038(X_2) + 0.018(X_3) + -0.017(X_4)$

(X_1 = education, X_2 = children, X_3 = age, X_4 = hours worked per week)

Holding all the other independent variables constant,

For each year of increase in education, television viewing should decrease by 0.115 hours.

For each additional child, television viewing decreases by 0.038 hours.

For each additional year of age, television viewing increases by 0.018 hours.

For each additional hour of work, television viewing decreases by 0.017 hours.

b.

Education, 0.202

Hours worked last week, –0.148

Age, 0.139

Number of children, –0.034

c. Together these four independent variables reduce the error in predicting TVHOURS by 8.3%. This is a weak prediction model.