

2 CHAPTER 1 SPSS Exercises > Explore additional data sets

<i>Data for Exercise</i>	
<i>Data Set</i>	<i>Description</i>
2012 states data.sav	This data set compiles official statistics from various official sources, such as the census, health department records, and police departments. It includes basic demographic data, crime rates, and incidence rates for various illnesses and infant mortality for entire states.
<i>Variables for Exercise</i>	
<i>Variable Name</i>	<i>Description</i>
State	The state from which data were collected
Murdercat	An indicator for the rate at which murders occur
Perindpoverty	The proportion of individuals below the poverty line
BurglaryRt	An indicator for the rate at which burglaries occur
MedianIncome	The median income for a state

1. First, take a look at the description and variable information in variable view in the 2012 states data set. What is the unit of analysis for these data?
2. In variable view, find the following variables and look at the variable and value labels. For each variable, identify (1) whether it is a quantitative or qualitative measure, (2) its level of measurement, and (3) if appropriate, the unit that it is measured in:
 - a. State
 - b. Murdercat
 - c. Perindpoverty
 - d. BurglaryRt
 - e. MedianIncome
3. In data view, find Alabama and Alaska (rows 1 and 2). What is the burglary rate for each state? Can you think of any reason for the differences between the two states?
4. **Creating a frequency table in SPSS:** We want to create a frequency table for the variable “Murdercat.” To do this, select analyze->descriptives->frequencies and put “Murdercat” in the box on the right. Looking at the output window, what proportion of states have 0–3 murders per 100,000? What proportion have fewer than 9 murders per 100,000?
 - a. **Sorting data in SPSS:** Let’s identify which states have the highest murder rate by sorting the “Murdercat” variable. To do this, go into data view and scroll until you find the variable Murdercat. Right click on the variable name for Murdercat and select sort descending, which will sort the data so that high values are on top. You can also sort cases by selecting data->sort cases and then placing Murdercat in the box. Now scroll left. What states have the highest murder rates?

2 CHAPTER 2 SPSS Exercises > Explore additional data sets

Data for Exercise	
Data Set	Description
NCVS lone offender assaults 1992 to 2013.sav	These data are from the National Crime Victimization Survey (NCVS), a nationally representative study of individuals' experience of criminal victimization. This particular data set contains responses from 1992 to 2013, allowing for larger numbers of uncommon offenses to be used in analyses. It also only includes data from respondents who reported either violent or nonviolent assaults by a single offender.
Variables for Exercise	
Variable Name	Description
Year	The year in which the data were collected. Ranges from 1992 to 2013.
Maleoff	The sex of the offender where 1 = male and 0 = female.
Injury	A binary variable indicating whether the respondent was physically injured during the assault where 1 = injured and 0 = uninjured.
Age_r	The age (in years) at which victimization occurred. This variable is cut off at 75.
V2129	This is a three-category nominal indicator of where a person lives. This is based on metropolitan statistical area (MSA) classifications, but loosely speaking, 1 = lives in city, 2 = lives near city (e.g., suburbs and outskirts), 3 = lives in a rural area away from cities.

- Creating Charts in SPSS:** Identify the level of measurement and which type of graph (histogram, pie chart, or bar chart) is appropriate for each of the following variables. If it is helpful, feel free to make frequency tables to determine this. Then create that chart in SPSS. To create the chart, select graphs->legacy dialogues->select bar, pie, or histogram. Then put the appropriate variable in the "category axis" or "variable" box:
 - Injury
 - Maleoff
 - Age_r
 - For this variable, also answer the following question(s): Is this variable's distribution skewed? If so, in what direction? Is there evidence of any extreme outliers?
 - V2129
- Pick two of the variables and/or charts from question 1. Write an explanation for your results in a way that would be easily understood by a layperson who isn't familiar with statistics. What does it tell us about people who are assault victims?
- What proportion of individuals were assaulted before they were 18? One way to see this is to look at a cumulative percentage table for the variable "Age_r." Create a frequency table for age_r, but this time focus on the far right column. What proportion of individuals were younger than 18 when they were victimized?
- The downside of the approach in question 3 is that you won't be able to easily make a graph of that finding. In this question, we'll do a *recode* to allow for a graph of the "Age_r" variable to make it easier to graph. Recoding typically involves grouping values of a variable in a way that simplifies them or makes useful distinctions:
 - Recoding Variables in SPSS:** Start with a plan. We want 0 to correspond to "younger than 18" and 1 to correspond to "18 and older."

- b. To begin the recoding process, select data->recode into *different* variables. Make sure you *don't* select “recode into same variables,” which will overwrite your original variable!
 - c. Put age_r in the box. Then, under the field for “output variable,” type the name “Vic18andover.” Click “change.”
 - d. Select “old and new values.” A new window will pop up. On the left side, you enter your original values; on the right side, you specify a value that they will be recoded to. First, select “system or user missing” and select “system-missing” in the new value; then press “add.” *This is an essential habit to get into as it will prevent you from unintentionally recoding missing values, which are treated as positive infinity by SPSS.*
 - e. Use the “greater than” and “less than” options to tell SPSS the values that you want to use as a cutoff; in this case, all values greater than 18 = 1, all values 17 or lower = 0. Be sure to hit “add” after each old and new value is specified.
 - f. The new variable will appear at the bottom of variable view mode. Under the “values” column, select the cell for “Vic18andover” and click the “...” option. Then copy the meaning for 0 and 1 specified in part 5a, selecting “add” for each one. This will make it so that output will show you what the categories you’ve created mean rather than just showing the numbers 0 and 1.
 - g. Create a frequency table of “Vic18andover.” Make sure that the proportion of individuals who were victimized younger than age 18 is the same as your answer for question 3. If it is not, double check your recode because you’ve made a mistake!
 - h. Create a bar chart of your new variable “Vic18andover.”
5. The NCVS has been conducted for many years using roughly the same measures. As such, it is an excellent tool for tracking trends over time. To begin with, let’s look at this theory: Some criminologists have suggested that as women enter the workforce and adopt more traditionally masculine roles, they will also engage in other more masculine behaviors. In particular, they will be more likely to commit crimes!
- a. To test this assertion, we’ll look at whether assault victims report a male or female offender at different rates over time. To do this, select graphs->legacy dialogues->line->simple. Then put the variable “Year” in the “category axis” box. In the “line represents” area, select “other statistic” and add the variable “Maleoff” into the box that lights up. Then select “ok”:
 - i. What is the general trend over time? The y axis can be multiplied by 100 to get the percentage of offenders who are male.
 - ii. Do these results support the theory presented?
 - iii. Look carefully at the scale of your graph. How large are these changes that have occurred over time?

2 CHAPTER 3 SPSS Exercises [➤ Explore additional data sets](#)

Data for Exercise	
Data Set	Description
Youth.sav	These data are from a random sample of students from schools in a southern state. Although not representative of the United States, the data set covers a variety of important delinquent behaviors and peer influences.

<i>Variables for Exercise</i>	
<i>Variable Name</i>	<i>Description</i>
V77	A five-category ordinal measure asking respondents how wrong they think their friends think it is to steal. Responses range from 1 (always wrong) to 5 (never wrong).
V79	A five-category ordinal measure asking respondents how wrong they think their friends think it is to drink. Response options are the same as for V77.
Certain	A scale indicating how likely respondents think it is that they will be punished for delinquent behavior. High values indicate high certainty.
Delinquency	A scale indicating the number of delinquent acts that respondents report participating in.
Gender	The gender of the respondents where 1 = male and 0 = female.
Parnt2	A scale indicating the amount of parental supervision that respondents receive, with values ranging from 2 to 8 where high values indicate more supervision.

- Use SPSS to construct a frequency table for the variable “Parnt2.” Based on that frequency table, derive the following with just the output. Any calculations should be done by hand with a calculator:
 - The mean parental supervision score
 - The median parental supervision score
 - The modal parental supervision score
 - For each of these points, write a sentence interpreting the value you receive in a way that a layperson would understand.
 - Of these three measures of central tendency, which do you think is most appropriate for describing this distribution and why? Be sure to consider whether skewness or outliers are a problem in your decision.
- Measures of Central Tendency in SPSS:** To do this, start off just like you were making a frequency table, but this time click “statistics.” A new menu will pop up, giving you lots of options. At the top right will be options for the mean, median, and mode; tick off all three. Have SPSS output the mean, median, and mode for the following variables:

 - V77
 - V79
 - Certain
 - Gender
 - First, what is the level of measurement for each variable? For each variable, identify which measures of central tendencies may be most appropriate.
 - For the interval- and ratio-level variables of this list, use the mean and median to describe the variables’ distribution. Does it appear to be skewed? If so, in what direction?
 - Pick one measure of central tendency for each variable. Explain why you prefer this measure for this variable. Then, write a sentence interpreting the measure.
- Create a histogram of the variable “Delinquency.” Looking at this distribution, what measure of central tendency will be most appropriate and informative? Will any be misleading? Use SPSS to calculate the appropriate measures of central tendency for this variable.
- Gender is coded 0 = female and 1 = male. This specific coding scheme is called “dummy coding,” and it is extremely common. Because it is both a binary variable and coded this way, the mean is actually still useful! It’s your job to find out what this mean represents in this specific circumstance:

- a. Construct a frequency for the variable “Gender.” Make note of the proportions that are male and female.
- b. Take a look at the mean from question 2 or the mean “gender” here. Now consider that the value .47 is rounded; the actual value is .465409.
- c. What does the mean of a dummy-coded variable actually tell us?

CHAPTER 4 SPSS Exercises ➤ Explore additional data sets

<i>Data for Exercise</i>	
<i>Data Set</i>	<i>Description</i>
Youth.sav	These data are from a random sample of students from schools in a southern state. Although not representative of the United States, they cover a variety of important delinquent behaviors and peer influences.
<i>Variables for Exercise</i>	
<i>Variable Name</i>	<i>Description</i>
V77	A five-category ordinal measure asking respondents how wrong they think their friends think it is to steal. Responses range from 1 (always wrong) to 5 (never wrong).
V79	A five-category ordinal measure asking respondents how wrong they think their friends think it is to drink. Responses are the same as for V77.
Certain	A scale indicating how likely respondents feel it is that they will be punished for delinquent behavior. High values indicate high certainty.
Delinquency	A scale indicating the number of delinquent acts that respondents report participating in.
Parnt2	A scale indicating the amount of parental supervision that the respondents receive, with high values indicating more supervision.
Moral_Cat	A categorical variable based on a scale that measured whether respondents thought delinquency was morally wrong. Five response categories range from 1 = “1 through 4” to 5 = “17 through 20”.

1. One measure of dispersion covered in this chapter is the variation ratio, which reflects the proportion of cases that fall outside the modal category. For each variable that follows, construct a frequency table and then calculate the variation ratio by hand. Then write a sentence or two explaining what that variation ratio tells us about the distribution of the variable:
 - a. V77
 - b. V79
2. Use frequency tables to calculate the range for the following variables:
 - a. Delinquency
 - b. V79
3. Use frequency tables for the variable “Parnt2” to calculate the variance and standard deviation for this measure. The mean for this measure is 6.23.
4. Using a frequency table and the procedures outlined for calculating the variance and standard deviation for grouped data, calculate the standard deviation and variance for the variable “Moral_cat.”

5. **Getting Measures of Dispersion in SPSS:** These can be obtained under the “frequencies” menu. Select analyze->descriptives->frequencies, and then click “statistics.” At the bottom left, you will see a box labeled “dispersion.” Check off “range” and “standard deviation”:
- What is the standard deviation and range for the following variables?
 - Delinquency
 - Certain

2 CHAPTER 5 SPSS Exercises > Explore additional data sets

<i>Data for Exercise</i>	
<i>Data Set</i>	<i>Description</i>
NCVS lone offender assaults 1992 to 2013.sav	These data are from the National Crime Victimization Survey (NCVS), a nationally representative study of individuals' experience of criminal victimization. This particular data set contains responses from 1992 to 2013, allowing for larger numbers of uncommon offenses to be used in analyses. It also includes only data from respondents who reported either a violent or nonviolent assault by a single offender.
2012 states data.sav	This state-level data set compiles official statistics from various official sources such as the census, health department records, and police departments. It includes basic demographic data, crime rates, and incidence rates for various illnesses and infant mortality for entire states.
<i>Variables for Exercise</i>	
<i>Variable Name (dataset)</i>	<i>Description</i>
Maleoff (NCVS)	The sex of the offender where 1 = male and 0 = female.
Victimreported (NCVS)	A binary variable indicating whether respondents reported their victimization to the police.
Relationship (NCVS)	A four-category nominal measure that tells us the victim's relationship to the offender. 0 = total stranger, 1 = slightly known, 2 = casual acquaintance, 3 = well known.
V3018 (NCVS)	A respondent's sex where 1 = male and 2 = female.
MurderRt (States)	Number of murders in a state per 100,000.
BurglaryRt (States)	Number of burglaries in a state per 100,000.

- Do victims know their assailants, or is it mostly strangers? Estimate a frequency table for the variable “Relationship” and use the output to find out! Use this output to do the following:
 - Use the restricted addition rule to calculate the following probabilities:
 - Pr(stranger)
 - Pr(well known)
 - Pr(slightly known, casual acquaintance, OR well known)
 - Use the probabilities you calculated to calculate the odds of each event occurring.
 - Write a sentence or two answering the question at the beginning of this question.

2. One of the most common methods for comparing two variables is to construct a cross-tabulation. We'll make one of these that compares victim gender with their relationship to the assailant to practice using the general addition rule of probabilities. To do this, go to analyze->descriptive statistics->crosstabs. Then, put gender in the row box and relationship in the columns.
 - a. Using this table, use the general addition rule to calculate the following probabilities:
 - i. Pr(female OR stranger)
 - ii. Pr(male OR well known)
 - iii. Pr(male OR well known)
 - b. Now use the multiplication rule to calculate predicted probabilities—recall that this notation refers specifically to *conditional probabilities*:
 - i. Pr(stranger | female)
 - ii. Pr(stranger | male)
 - iii. Pr(well known | female)
 - iv. Pr(well known | male)
 - v. Zoom out for a moment and think substantively about these results:
 1. Which gender has a greater probability of being assaulted by a stranger? By a well-known acquaintance?
 2. Do women have a greater probability of being assaulted by a stranger or a well-known acquaintance?
3. Hypothesis tests are quick and easy in SPSS. Let's do a hypothesis test to answer the following research question: Are women more likely to assault another person than men?
 - a. First, write out the null and alternative hypotheses. In this case, no sex differences would be evidenced by an even probability of assault. That is, $p = .50$.
 - b. What probability distribution will be used to answer this question?
 - c. What alpha level will you use?—typical values are .05, .01, and .001, with smaller alpha values reflecting more difficult criteria.
 - d. **Binomial Hypothesis Tests in SPSS:** Do this by clicking analyze->nonparametric tests->legacy dialogues->binomial. Put the variable "Maleoff" into the "test variable list." Set "test proportion" to .50:
 - i. The output you receive gives you two pieces of information: The "observed prob" is the probability of being in group 1 versus group 2. The "Exact Sig (2-tailed)" column will have a value in it. This is the exact value for p in this analysis; if it is equal to .000, it is actually even smaller than .001.
 - e. Using the output from this model, what do you conclude about the null hypothesis? Substantively, what does this mean?
4. Let's do another hypothesis test to answer the following research question: Are victims more likely to report their own victimization than to have it be reported by someone else or another avenue?
 - a. First, write out the null and alternative hypotheses. In this case, the probability that any victimization will be reported is .447.
 - b. What probability distribution will be used to answer this question?
 - c. What alpha level will you use?—typical values are .05, .01, and .001, with smaller alpha values reflecting more difficult criteria.
 - d. In SPSS, run the hypothesis test. Do this by clicking analyze->nonparametric tests->legacy dialogues->binomial. Put the variable "Victimreported" into the "test variable list." Set "test proportion" to .447.
 - e. Using the output from this model, what do you conclude about the null hypothesis?

- f. Substantively, what does this mean? Explain the results of your hypothesis test in a way that a layperson could understand easily.

The remaining questions are answered using the 2010 states data.sav.

5. What does the distribution of homicide look like throughout the United States?
 - a. First, create a histogram of the variable “MurderRt.” Describe the distribution. Is it relatively normally distributed or skewed?
 - b. Use the frequency menu to get the standard deviation of MurderRt. Use it to answer the following:
 - i. What range would you expect 68% of states’ murder rates to fall within?
 - ii. What range would you expect 95% of states’ murder rates to fall within?
 - iii. What range would you expect 99.7% of states’ murder rates to fall within?
6. Repeat question 5 using the variable “RobberyRt.”
7. The z scores are easily obtained in SPSS and can then be used to answer a wide array of questions. Let’s get z scores for the murder rates across states. To do this, go to analyze->descriptive statistics->descriptives. Then, put MurderRt in the variable box and check off the “save standardized values as variables” option. A new variable called “ZMurderRt” will show up at the bottom of variable view:
 - a. Look at the z score for Alabama. What proportion of states are expected to have a *higher* murder rate?
 - b. Look at the z score for Alaska. What proportion of states are expected to have a *lower* murder rate?
 - c. Now compare z scores for Alabama and Alaska. What proportion of states is expected to have a murder rate *between* that of Alaska and Alabama?
8. The standard error of the sampling distribution is easily attained in SPSS. The easiest way to do so is to go to the frequency menu and select the “statistics” menu. Then, tick off “S.E. Mean”:
 - a. What is the standard error for the murder rate (MurderRt) in this case?
 - b. What is the standard error for the burglary rate (BurglaryRt) in this data set?

CHAPTER 6 SPSS Exercises > Explore additional data sets

<i>Data for Exercise</i>	
<i>Data Set</i>	<i>Description</i>
2012 states data.sav	This data set compiles official statistics from various official sources such as the census, health department records, and police departments. It includes basic demographic data, crime rates, and incidence rates for various illnesses and infant mortality.
<i>Variables for Exercise</i>	
<i>Variable Name</i>	<i>Description</i>
State	ID for state that the data are from.
AssaultRt	Number of assaults in a state per 100,000.
RobberyRt	Number of robberies in a state per 100,000.
MVTheftRt	Number of motor vehicle thefts in a state per 100,000.
Assault_bin	A binary recode of the assault rate where 0 = less than 200 assaults per 100,000 and 1 = 200 or more assaults per 100,000.

1. First, take a look at the 2010 state data set. Answer these questions:
 - a. How many cases are in this data set?
 - b. What does the variable “State” tell you if you look at a frequency?
 - c. What unit of analysis should be used for these data?
 - d. Assume that official records are accurately collected. Are these sample data or population data?
 - e. Say we took an average of the murder rates in these data. What will that average tell us?
2. Take a look at the average assault rate for the United States by looking at descriptives for the variable AssaultRt (analyze->descriptive statistics->frequencies->statistics):
 - a. What do you find? Hang on to that mean!
 - b. Consider the type of data you’re working with based on your answer to question 1. What does the standard deviation you’ve been given by SPSS reflect? Recall Table 4.12 in Chapter 4. What symbol is appropriate for this value?
3. If you decided in question 1 that these are *population*-level data, you’d be correct! Such data are rather rare, but we can use them to explore how close estimates from a *sample* from the population come to capturing the actual population mean. We’ll do this by taking a random sample of states, using it to generalize to the population as a whole with a confidence interval:
 - a. **Taking a random sample:** To select 25 states randomly, select data->select cases and then click “sample” under the “random sample of cases” option. In the pop-up menu, tell it to select 25 of 50 states randomly. Some rows should now be crossed off in data view.
 - b. Get the mean and standard deviation for this subsample. Use this information to calculate a standard error and 95% confidence interval for the variable AssaultRt:
 - i. What is the appropriate *t* statistic for this confidence level?
 - c. Interpret the confidence interval in technical terms in a sentence. Then write another sentence that a layperson would be able to understand easily.
 - d. Return to your results from problem 2. Is your confidence interval able to capture the true population mean?
4. **Estimating a confidence interval in SPSS:** To get SPSS to give you a confidence interval, you must select analyze->descriptives->explore. For this case, we want to estimate value for the population mean robbery rate from our sample. The output you’re interested in is the mean and the confidence interval for the mean:
 - a. First, get the population mean for the robbery rate (RobberyRt)—all 50 states should be selected.
 - b. Now, take a random sample of 25 states. Estimate a 95% confidence interval around the mean:
 - i. Calculate the standard error of the mean by hand using the information in this output.
 - ii. Write a sentence interpreting this confidence interval so that a layperson could understand it.
 - iii. Is the true population mean captured in this confidence interval? Why or why not?
 - c. Under the “explore” menu, click “statistics.” Change the confidence interval for the mean to 90%:
 - i. How has the confidence interval changed compared with part 4b? Why has this change happened?
 - ii. Is the true population mean captured in this confidence interval? Why or why not?
 - d. Under the “explore” menu, click “statistics.” Change the confidence interval for the mean to 99%:
 - i. How has the confidence interval changed compared with parts 4b and 4c? Why has this change happened?
 - ii. Is the true population mean captured in this confidence interval? Why or why not?
 - e. Repeat parts 4b through 4d with a random sample of 10 states:
 - i. How has the standard error of the mean changed? Why has this happened?
 - ii. Explain the relationship among sample size, level of confidence, and the confidence interval width.

5. Repeat parts 4a to 4c with the variable MVTheftRt in a subsample of both 10 and 25 states.
6. SPSS does not allow for the automatic estimation of confidence intervals around a proportion. We can easily acquire the information needed to create this estimate by hand from a frequency table, however. Let's build a confidence interval around an estimate for the proportion of states with an assault rate higher than 200 assaults per 100,000:
 - a. Create a frequency for the variable Assault_bin to get the population proportion; make sure you aren't looking at a random sample! Remember that this value is presented as a percentage in SPSS. You'll need to divide it by 100 to convert it to a proportion.
 - b. Create another frequency table for the same variable, this time using a random subsample of only 25 states.
 - c. Use the information provided in this table to calculate the standard error of the proportion and both 95% and 99% confidence intervals; you will need to look in the back of the book for the appropriate t statistic values.
 - d. For each confidence interval you estimate, write a sentence interpreting the result in a way that a layperson would be able to understand.
 - e. Do your confidence intervals capture the true population proportion? Why or why not?

2 CHAPTER 7 SPSS Exercises > Explore additional data sets

<i>Data for Exercise</i>	
<i>Data Set</i>	<i>Description</i>
Youth.sav	These data are from a random sample of students from schools in a southern state. While not representative of the United States, they cover a variety of important delinquent behaviors and peer influences.
<i>Variables for Exercise</i>	
<i>Variable Name</i>	<i>Description</i>
Certain	A scale indicating how likely respondents feel it is that they will be punished for delinquent behavior. High values indicate high certainty.
Delinquency	A scale indicating the number of delinquent acts that respondents report participating in.
Moral	A scale that measures if respondents think delinquency is morally wrong. High values indicate that delinquency is viewed as morally wrong.
Parnt2	A scale indicating the amount of parental supervision that respondents receive, with values ranging from 2 to 8 and high values indicating more supervision.

1. **A fictional example with real data:** Imagine that you are a school official working to reduce delinquency within the student body. You and your staff members implemented a wide array of programs trying to fix your delinquency problem:
 - a. A meeting was held explaining to students that it was very likely that they would be caught if they were bad.
 - b. You held a session explaining why delinquent behaviors were harmful to people other than themselves and, therefore, were wrong.
 - c. You contacted students' parents and told them it was very important that they monitor their children's behavior and engagement in school.
2. You measured student delinquency and student attitudes about deviance both before and after the intervention. To begin with, you want to see if the students view punishment as being more certain than the original test, which had a mean of 10.40:

- a. Formally write your statistical hypotheses.
 - b. Use the frequencies dialog menu to get the standard deviation, mean, and sample size for the variable “Certain.”
 - c. What is the appropriate test statistic and sampling distribution for this case? Will you do a one-tailed or a two-tailed test? Why?
 - i. What is the appropriate value for the test statistic for your one/two-tailed test when:
 1. $\alpha = .05$
 2. $\alpha = .01$
 3. $\alpha = .001$
 - d. By using the mean and standard deviation, calculate your test statistic by hand; recall that your comparison “population” value is 10.40.
 - e. What do you conclude about your hypotheses? Is the conclusion the same at all levels of significance?
 - i. If you ran a one-tailed test, compare with a two-tailed test. Likewise, if you started with a two-tailed test, compare with a one-tailed test. Are your findings the same?
 - f. Write a few sentences summarizing these results to your boss. Be sure to include mention of the type of test you used, what the result was, and the highest level at which the results were significant, if at all. Also write a sentence about the “so what” of your results: What does this mean for your intervention, and how large was the change produced?
3. You quickly realize that doing these tests by hand will take quite some time. Fortunately, SPSS allows us to conduct z tests very quickly, although the option is labeled as a t test. In reality, this function does both; t and z statistics are equivalent when your sample is greater than 120. With this knowledge in mind, you move on to the possibility that students view crime as more wrong following the intervention, a key element of social bond theories. The original survey of students found that they had a mean score of 16.5:
- a. Formally write your statistical hypotheses.
 - b. What is the appropriate test statistic and sampling distribution for this case? In this case, assume a two-tailed test will be used.
 - c. Determine an alpha level that you will use. You do not need to find the critical value.
 - d. Use SPSS to perform a one-sample z or t test. Select analyze->compare means->one sample t test. Put the variable “Moral” in the “test variable” box, and set your test value to the comparison mean, 15.2:
 - i. The output SPSS provides several key pieces of information, most of which are intuitively labeled. In this case, the “ t ” statistic provided is actually a z statistic due to your sample size. The main box to notice is labeled “sig (2-tailed),” which contains the exact p value, which is compared against your specified alpha. A value of .000 means your p value is actually something like .000001—certainly significant at $p = .001$.
 - e. What do you conclude about your hypotheses? Is the conclusion the same at all levels of significance?
 - f. Write a few sentences summarizing these results to your boss. Be sure to include mention of the type of test you used, what the result was, and the highest level at which the results were significant, if at all. Also write a sentence about the “so what” of your results: What does this mean for your intervention, and how large was the change produced?
4. You receive a call with some very bad news. Your research assistant jumped the gun and destroyed a lot of your surveys before they had been fully entered! At the end of the day, you now have only 7% of the data that you thought you had for the delinquency and parent supervision measures. You decide to push on with what you’ve got—only 81 responses instead of 1,272! Your next task is to see if parents were more engaged in supervising their children than in the original test, which has a mean score of 6.1. Your research assistant also broke SPSS, so you’ll have to do this one by hand:

- a. Select a random sample of 7% of your data by selecting data->select cases and selecting “sample” and “approximately 7% of cases.”
 - b. Formally write your statistical hypotheses.
 - c. Use the frequencies dialogue menu to get the standard deviation, mean, and sample size for the variable “Parnt2.”
 - d. What is the appropriate test statistic and sampling distribution for this case? Will you do a one-tailed or two-tailed test? Why?
 - i. How many degrees of freedom do you have?
 - ii. What is the appropriate value for the test statistic for your one/two-tailed test when:
 1. $\alpha = .05$
 2. $\alpha = .01$
 3. $\alpha = .001$
 - e. By using the mean and standard deviation, calculate your test statistic by hand; recall that your comparison “population” value is 6.1.
 - f. What do you conclude about your hypotheses? Is the conclusion the same at all levels of significance?
 - i. If you ran a one-tailed test, now compare with a two-tailed test. Likewise, if you started with a two-tailed test, compare with a one-tailed test. Are your findings the same?
 - g. Write a few sentences summarizing these results to your boss. Be sure to include mention of the type of test you used, what the result was, and the highest level at which the results were significant, if at all. Also write a sentence about the “so what” of your results: What does this mean for your intervention, and how large was the change produced?
5. You are finally ready to move on to your final outcome: whether students are engaging in less delinquent behaviors. SPSS is working again, so we’ll do this analysis with software. Before the interventions, you found students reported a mean of 38 delinquent activities:
- a. Select a random sample of 7% of your data (see question 3a).
 - b. Formally write your statistical hypotheses.
 - c. What is the appropriate test statistic and sampling distribution for this case? In this case, assume a two-tailed test will be used.
 - d. Determine an alpha level that you will use. You do not need to find the critical value.
 - e. Use SPSS to perform the hypothesis test. Do this by selecting analyze->compare means->one sample *t* test. This process and associated output is explained in question 2d, and this time your comparison value is 38.
 - f. What do you conclude about your hypotheses? Is the conclusion the same at all levels of significance?
 - g. Write a few sentences summarizing these results to your boss. Be sure to include mention of the type of test you used, what the result was, and the highest level at which the results were significant, if at all. Also write a sentence about the “so what” of your results: What does this mean for your intervention, and how large was the change produced?
6. Step back and look at all the results from questions 1 through 4. If you had to advocate for the use of just *one* intervention, which one seems to be most effective, and why?

2 CHAPTER 8 SPSS Exercises > Explore additional data sets

<i>Data for Exercise</i>	
<i>Data Set</i>	<i>Description</i>
NCVS lone offender assaults 1992 to 2013.sav	These data are from the National Crime Victimization Survey (NCVS), a nationally representative study of an individual's experience of criminal victimization. This particular data set contains responses from 1992 to 2013, allowing for larger numbers of uncommon offenses to be used in analyses. It also includes only data from respondents who reported either violent or nonviolent assault by a single offender.
<i>Variables for Exercise</i>	
<i>Variable Name</i>	<i>Description</i>
V2129	This is a three-category indicator of where a respondent lives. This is based on metropolitan statistical area (MSA) classifications, but loosely speaking, 1 = lives in city, 2 = lives near city (suburbs, outskirts), 3 = lives in a rural area away from city.
V3023	The race of the respondent where 1 = White, 2 = Black, 3 = Native American, 4 = Asian.
V3018	A respondent's sex where 1 = male, 2 = female.
Maleoff	The sex of the offender (attacker) where 1 = male, 2 = female.
Relationship	A four-category nominal measure that tells us the victim's relationship to the offender where 0 = total stranger, 1 = slightly known, 2 = casual acquaintance, 3 = well known.

- Are people from different races more likely to be assaulted if they live inside or outside of the city? For instance, are Whites more likely to be assaulted in the city than non-Whites? We can examine this question with a contingency table in SPSS and then calculate the appropriate cell proportions:
 - In this case, what is your independent variable?
 - Constructing a contingency table (i.e., cross-tabulation) in SPSS: First, select analyze->descriptives->crosstabs. Place your two variables, "V2129" and "V3023," into the row and column boxes. Only put one variable per box; pay close attention to whether the independent variable is placed along rows or columns.
 - Calculate cell percentages for this table, using the appropriate marginal to allow for comparisons across the categories of your independent variable.
 - Looking at the proportions you've calculated, what do your results tell you about the relationship between geographic region and race of victim? What is the relative risk of a Black person being assaulted in the city versus a rural area (not MSA)?
 - Can you think of any reasons as to why you've found what you did; why might this crime be unequally distributed by geography?
- Who hurts whom? Are men (or women) more likely to assault a member of a different sex or the same sex?
 - Write a hypothesis about this question; this does not need to be a formal hypothesis yet.
 - First, we must set up our data so we are looking only at assaults that result in injury. To do this, we will run analyses on only individuals who reported being injured. Under data->select cases, click the "if" button. Enter the expression "injured = 1" into the field. You will know this worked if you calculate a frequency of the variable injured and see that it always equals 1.
 - Getting cell percentages from SPSS:** First, select analyze->descriptives->crosstabs, as usual. Then, put one variable in the row slot and the other variable in the column slot; your variables are named "Maleoff" and

- “V3018” this time. Then, click the “cells” button and look for the “percentages” section of the menu that pops up. If your independent variable was in the “rows” box, tick off the “rows” option, and if your independent variable was in the “columns” box, tick off the “columns” option. This makes it so that percentages adding up along either rows or columns are provided, allowing for easy comparisons of categories of the independent variables.
- d. Interpret the results of the cross-tabulation fully. What are the implications of these results for our understanding of victimization?
3. Using the cross-tabulation from question 2, calculate the chi-square test of independence:
 - a. Formally state your research hypothesis, statistical distribution, alpha level, and critical values.
 - b. What is your χ^2 test statistic?
 - c. What do you conclude about your hypotheses?
 4. Now let’s assess the strength of the association from question 3: Calculate the appropriate measure of association by hand.
 - a. What do these statistics tell us? Write a sentence or two assessing the strength of the association.
 - b. Take another look at your results from questions 2 through 4. Consider the following assertion: “Women inflict violence against men at the same rate as men inflict violence against women.” Based on all of your results, defend or critique this statement. Be sure to discuss the results of your hypothesis test and the strength of the association that you found.
 5. One common idea in our culture is that we need to be wary of strangers as we never know when one may actually be a rapist, an attacker, or a robber. On the other hand, people we know are characterized as safe, trustworthy, and predictable. This idea was touched on in Chapter 5, but now let’s expand on those results and add hypothesis tests. Put briefly, the idea of “stranger danger” would lead us to hypothesize that most attackers in a violent crime are unknown:
 - a. If you haven’t done so already, select only assault cases that resulted in injury to the victim as described in question 2b.
 - b. Construct a frequency table for the variable “Relationship.” What does this tell us about the perpetrators in violent assaults?
 - c. Perhaps strangers are a particular threat to just women as compared with men. To investigate this, do the following:
 - i. Create a cross-tab for the variables “V3018” and “Relationship.” Be sure to put your independent variable in the “column” box and to select the column percentage option in the cell menu.
 - ii. Conduct a chi-square hypothesis test in SPSS. To do this, go to the cross-tab dialog box and then select “statistics.” Tick off “chi square”:
 1. The output provided gives the chi-square test statistic and its associated p value (labeled “sig”). For now, focus only on the “Pearson’s Chi Square” row.
 2. Write a few sentences interpreting the hypothesis test; be sure to note your alpha level, the value of your chi-square statistic, and your conclusions about the null hypothesis.
 - iii. **Using SPSS to estimate categorical measures of association:** In this case, use Cramer’s V . This can be found under the same menu where you located the chi-square test:
 1. Write a few sentences explaining what this result suggests about the strength of the association.
 - d. Based on these analyses, what would you conclude about the idea of stranger danger in the case of violent assault? Is it accurate to caution that strangers should be feared and that well-known acquaintances are safe? Be sure to use all the details from your analysis to back up your argument.

CHAPTER 9 SPSS Exercises > Explore additional data sets

<i>Data for Exercise</i>	
<i>Data Set</i>	<i>Description</i>
Youth.sav	These data are from a random sample of students from schools in a southern state. Although not representative of the United States, it covers a variety of important delinquent behaviors and peer influences.
<i>Variables for Exercise</i>	
<i>Variable Name</i>	<i>Description</i>
D1	A binary variable based on the number of delinquent acts a respondent reported. A 0 indicates that the respondent reported 1 or fewer acts, whereas a 1 indicates 2 or more acts.
Lowcertain_bin	Binary indicator of whether respondents felt there was certainty that they'd be punished for delinquent behaviors. 1 = low certainty, 0 = high certainty.
Gender	The gender of the respondent where 1 = male and 0 = female.
V2	Age of the respondent in years.
Moral	A scale that measures whether respondents thought delinquency was morally wrong. High values indicate that delinquency is viewed as morally wrong.
Delinquency	A scale indicating the number of delinquent acts that an individual reports participating in.
Studyhard	Binary indicator of studying behavior of respondents where 0 = studies less than 8 hours per week and 1 = studies 8 or more hours per week.

1. Are delinquent teens more likely to be older than teens who don't report delinquency? One way to look at this question is with a t test. In this case, we'll compare the mean age for delinquent students with the mean age of non-delinquent students:
 - a. State research and null hypotheses for this topic.
 - b. What type of test should you use: an independent-samples t test or a matched-group test?
 - c. For this analysis, use an alpha of .001 for your significance level.
 - d. **Conducting an independent-samples t test in SPSS:** To run a t test in SPSS, select analyze->compare means->independent-samples t -test. Put the variable "V2" (age) in the "test variable" box. Click "define groups" and enter 0 in one box and 1 in the other. Then, put our grouping variable, "D1," in the "grouping variable" box. Select Ok:
 - i. The output for this needs some explanation. The first box that comes out tells you the number of cases in each group (n) and the mean and standard deviation for each group. The second box has two rows that correspond to t tests conducted with equal or unequal variances assumed. Levene's test for equality of variances is an F test comparing group variances that will be addressed in a later question. You are interested in the columns labeled "T" and "sig. 2-tailed"; these are your test statistic and its exact p value. Also of interest is the "mean difference" box, which contains the value for the mean of group 1 minus the mean of group 2.
 - e. Assume that a difference of .15 in standard deviation is evidence that you have unequal variances. After looking at the standard deviations for either group, do you conclude that these groups have equal or unequal variances?
 - f. By using the appropriate variance assumptions, what do you conclude about the null hypothesis?
 - g. Write a sentence explaining what this result tells us about delinquency in substantive terms.

2. Are students who spend lots of time studying likely to also view delinquency as wrong? Some theorists would suggest this should be the case as high adherence to social roles (e.g., student, good kid) means that the penalties for delinquency will have a higher social cost. Let's investigate with a t test:
 - a. State research and null hypotheses for this topic.
 - b. What type of test should you use: an independent-samples t test or a matched-group test?
 - c. For this analysis, use alphas of both of .05 and .001 for your significance level.
 - d. Conduct the independent-samples t test in SPSS (described earlier). In this case, use the variable "Moral" as the dependent variable and "Studyhard" as the independent variable.
 - e. **Levene's test for equality of variances:** The easiest way to determine whether we should assume equal variances in our t test is to use an F test comparing the variances of both groups. The two columns for Levene's test for equality of variances do just that. The output shows an F test and its associated p value. The null hypothesis in this case is that variances are equal, whereas the research hypothesis is that they are unequal. In other words, a significant test means we should use the "equal variances not assumed" test row:
 - i. What is your decision about the null hypothesis in Levene's test? What assumptions should you use for your t test?
 - f. What is your conclusion regarding the null and research hypotheses?
 - g. Substantively, do these results support the theory described earlier, or do they contradict it? Use your data to explain why.
3. Do boys commit more delinquent acts, on average, than girls? Perform an independent-samples t test with the variables "Gender" and "Delinquency" to test this question. Use questions 1 and 2 to guide you through this process; be sure to check your F test in determining the appropriate test to use. After the test, provide a criminological or sociological explanation for why you found what you did.
4. Just before being hauled off to jail, TV show villains often say, "I would have gotten away with it if it wasn't for you meddling kids!" The underlying statement is that they committed the crime because they thought they could get away with it. We could ask, then, are individuals who think they won't get punished more likely to commit delinquent acts? Conduct an independent hypothesis test between the variables "Lowcertain_bin" and "Delinquency." Be sure to go through all steps of the hypothesis testing process. After the analysis, write a few sentences about the practical implications of this result.
5. "But I didn't know it was wrong!" is a common statement made by children when they have been caught doing something bad. Perhaps the belief that their behaviors just aren't so wrong might lead to more delinquency. To research this, conduct a t test of the variables "Moral" and "Delinquency." You may notice that the variable "Moral" is not binary; when defining groups in the t test menu, use the value 16 as a cut point. This will compare individuals with scores that are greater than or equal to 16 against those who score less than 16. What do you conclude about the null hypothesis? What is the practical significance of this result?

CHAPTER 10 SPSS Exercises ➤ Explore additional data sets

<i>Data for Exercise</i>	
<i>Data Set</i>	<i>Description</i>
Monitoring the future 2013 grade 10.sav	This data set contains variables from the 2013 Monitoring the Future (MTF) study. These data cover a national sample of 10th graders with a focus on monitoring substance use and abuse.


Variables for Exercise	
Variable Name	Description
V1070	A student's race where 1 = Black, 2 = White, and 3 = Hispanic.
SchoolSkip	A count of the number of days a student has skipped school or classes.
V7253	An ordinal measure asking whether the student has any friends who have dropped out of school, where 1 = none, 2 = a few, 3 = some, and 4 = most if not all.
LowParentEduc	A three-category summary of both parents' education where 0 = both parents have high school diplomas, 1 = one parent has less than a high school education, 2 = both parents have less than a high school education.
V7214	An ordinal measure asking students the number of hours they are home alone after school on a typical day. Responses range from 1 = 0 hours to 6 = 5 or more hours.
SmokeDrink	A continuous index measuring the quantity and frequency of smoking and drinking self-reported by respondents.

- Are students of different races more likely to skip out on school? Criminologists are interested in this question because school acts as a form of social control; if you're at school, you won't have as many chances to engage in crime or interact with the justice system! We can test this question using a one-way ANOVA in SPSS with the variables "SchoolSkip" and "V1070":
 - First, state the null and research hypotheses for this ANOVA.
 - Second, determine the critical value of F you will need for your results to be significant at $p < .01$. You will need to calculate your within-group and between-groups degrees of freedom in this case; in this case, $n = 10,216$ and $k = 3$.
 - Estimating an ANOVA model in SPSS:** Doing this by hand would take weeks, but SPSS does it in seconds. To do this, select analyze->compare means->one-way ANOVA. Place the variable "V1070" in the "factor" slot and "SchoolSkip" in the dependent list:
 - Output provided gives the sum of squares within and between groups, your degrees of freedom, and F statistics. It also provides the p value for your specific F statistic.
 - What is your F statistic? What do you conclude about the null hypothesis in this case?
- You may have noticed that an ANOVA test on its own isn't a very precise assessment of our research question. Let's follow this up by estimating the difference between pairs of means and the strength of the association:
 - Post Hoc Tests in ANOVA:** In the one-way ANOVA menu, you will notice an option to "post-hoc" tests. Click that button and select "Tukey," which will give you an estimation of Tukey's honest significant difference:
 - The output provided compares one racial category with all the others, providing the statistical significance of each difference. So, for instance, the first row compares Blacks against Whites and shows that the difference between the means for Blacks and Whites is .449 days skipped.
 - Interpret the remaining differences between racial groups. Are all groups significantly different from one another? Which group means are closest/furthest from one another?
 - What new information does this test add to our ANOVA interpretation?
 - By using the ANOVA sum of squares output, calculate η^2 by hand:
 - What new information does this statistic tell us?
 - Consider the difference between statistical and substantive significance. Do we have both in this case?
- Do students who hang out with individuals who drop out of school also tend to skip out on school themselves? Run a one-way ANOVA test with pairwise comparisons and estimate η^2 (by hand) using the variables "V7253" and "SchoolSkip":

- a. Conduct the hypothesis test fully, stating your null hypotheses and identifying test statistics ($n = 12,449$ and $k = 4$ in this case). What do you conclude about the null hypothesis?
 - b. What do the pairwise comparisons in this case tell you? What groups are most prone to skipping school?
 - c. What does η^2 tell us in this case? What proportion of the variability in school skipping can be explained by one's peer network?
 - d. By thinking substantively, what does this tell us about the role of peer networks? Do you think it might relate to other areas of deviant behavior as well?
4. Are kids who have lots of unsupervised time more likely to use alcohol or drugs than other kids? Use the variables "SmokeDrink" and "V7214" to test this hypothesis. Also estimate pairwise comparisons and η^2 :
 - a. Conduct the hypothesis test fully, stating your null hypotheses and identifying test statistics ($n = 12,169$ and $k = 6$ in this case). What do you conclude about the relationship?
 - b. What do the pairwise comparisons in this case tell you? Use an alpha of .001 for this test. Are any groups not statistically different? It may help to visually rank each group against one another in a diagram.
 - c. What does η^2 tell us in this case?
 - d. Given your results, what are your policy recommendations?
 5. Often researchers talk about the importance of socioeconomic status in producing heightened risk for criminal and delinquent activity. The education of a student's parents is one way to look at this difference. Let's see whether the variable "LowParentEduc" is associated with higher levels of student smoking and drinking:
 - a. Conduct a hypothesis test to examine whether there is a relationship between parent education and child alcohol and tobacco use. Fully state your null hypotheses and identifying test statistics ($n = 10,726$ and $k = 3$ in this case). What do you conclude about the relationship?
 - b. What do the pairwise comparisons in this case tell you? Use an alpha of .001 for this test. Are any groups not statistically different from one another?
 - c. What does η^2 tell us in this case?
 - d. Take a moment to consider the "why" part of this result; what mechanisms might explain why the results you have found might lead to delinquent behaviors?

CHAPTER 11 SPSS Exercises ➤ Explore additional data sets

<i>Data for Exercise</i>	
<i>Data Set</i>	<i>Description</i>
2012 states data.sav	This state-level data set compiles official statistics from various official sources such as the census, health department records, and police departments. It includes basic demographic data, crime rates, and incidence rates for various illnesses and infant mortality for entire states.
<i>Variables for Exercise</i>	
<i>Variable Name</i>	<i>Description</i>
TobaccoDeathRt	The tobacco-related death rate per 100,000 for a state.
PerIndPoverty	The proportion of individuals below the poverty line in a state.
BurglaryRt	The burglary rate reported per 100,000 for a state.
RobberyRt	The robbery rate reported per 100,000 for a state.

1. Your adviser, Dr. Smartypants, says, “Let’s see how state characteristics end up predicting the robbery rate. Consider the following question: Is there a link between the unhealthy behaviors people engage in and the crime rate? I’m just fishing for results here, but let’s see what we find!” He tells you to look into the relationship between the variables “TobaccoDeathRt” and “BurglaryRt” to see whether there is a relationship. He wants a graph of the two variables, estimates of the strength of the association, and a regression model to be fitted:
 - a. **Constructing a Scatterplot:** Before running the regression model, we should look at the relationship of the two variables using a scatterplot. This is done by selecting graphs->legacy dialogues->scatter/dot->simple scatter. Then, put your independent variable on the y axis and put the dependent variable on the x axis. Hang on to this graph.
 - b. **Estimating Pearson’s Correlation Coefficient:** To estimate Pearson’s r , select analyze->correlate->bivariate:
 - i. This output provides a matrix of correlations. Generally, you are interested in only the bottom left half as the top right half is redundant. Each cell contains several pieces of information: the correlation coefficient, the direction of the association, and the significance of the association presented in each cell. You can tell the level of statistical significance for a particular coefficient at a glance by the asterisks by each correlation coefficient where $*p < .05$, $**p < .01$, and $***p < .001$:
 1. Interpret the correlation coefficient, making special note to address the strength and direction of the association.
 - c. **Estimating a Regression Model:** To do this, select analyze->regression->linear and put the independent and dependent variables in the correct spots:
 - i. The output provided gives several pieces of information. The “model summary” box contains R -squared and R (but not its direction!) Notice that SPSS labels them both with capital letters. The ANOVA (analysis of variance) box is largely ignorable for your purposes. The “coefficients” box is what you are interested in. “Constant” refers to your intercept. The slope for your regression line is in the row named after your independent variable in the “unstandardized coefficients: B” column. Significance tests for the slope can be found on the far right under the “sig” box:
 1. Write the equation for your regression model in full.
 2. Interpret your regression coefficient. What does it tell us in this case?
 3. Are these results statistically significant? What are the hypotheses you are testing?
 4. Interpret the intercept as well. Is it useful in this scenario?
 5. Note the value for R -squared. What does this coefficient tell us, and why is it important?
 - d. Now, put together a few sentences summarizing all your results to Dr. Smartypants. How comfortable are you with these accuracy results? Can you detect any caveats or concerns that must be considered?
2. Dr. Smartypants pulls you aside and confesses that he is more a visual person and would like you to give him a graph with the regression line plotted on it directly:
 - a. **Adding a Regression Line to a Scatterplot:** First, construct a scatterplot as usual. Then, double click on the image of the graph plot. The Chart Editor will pop up. To add a regression line to the plot, simply select the  button.
3. You turn in your report to Dr. Smartypants, who is clearly very excited. He says, “Holy macaroni! This proves it! The smokers are also robbing everyone; smoking causes criminal impulses! I’ve got to publish this right away!”:
 - a. What do you think of his conclusion? Do you think it is an accurate interpretation of the data? Write a statement of support of his interpretation if you agree, or write a rebuttal and alternative interpretation if you disagree.

4. You decide that you want to be sure about these results before publishing them by checking for another plausible relationship; perhaps poverty is linked to a high robbery rate:
 - a. Run all the analyses that you did for part 1 again, this time using the variable “PerIndPoverty” as your independent variable.
 - b. Write a summary of your results for Dr. Smartypants that incorporates the results from your scatterplot, correlation, and regression analysis.
5. Based on the analyses you have conducted so far, what seems to be the most plausible explanation for robbery? If both variables are associated, why might that be the case and which one do you think is the root cause? Which explanation seems to be more theoretically sound? Are you in a position to make claims of causality?
6. Dr. Smartypants asks you for one more favor before you publish your paper together. He wants you to see whether the dependent variable, “BurglaryRt,” has any problems that might weaken the certainty of your conclusions. In essence, he wants you to check some of your regression assumptions. If you detect any violations of assumptions, consider how they may have influenced your conclusions; are your coefficients biased or just your hypothesis tests, for instance?:
 - a. What level of measurement is this variable?
 - b. Use SPSS’s graphing functions to construct a histogram of “BurglaryRt.” How would you describe this distribution—skewed left, skewed right, or normally distributed? Is there any evidence of extreme outliers?
 - c. Return to your scatterplot. Is there any evidence of a nonlinear relationship?
 - d. **Estimating Errors (residuals) in SPSS:** Under the regression dialogue, select “save,” and in the top right you’ll find a box for “residuals.” Tick off the “unstandardized” box. *Estimate a model with poverty predicting burglary.* A new variable, typically called “RES_1,” will be created. Create a histogram. Are your errors normally distributed with a mean of roughly 0? Is there any evidence of heteroscedasticity in your errors (e.g., skewness and sudden enormous spikes)?
 - e. Think carefully about the data you are looking at. They come from official police records. Can you think of anything that might threaten their accuracy?
7. If you would like additional practice with conducting regression analyses in SPSS, go through exercises 1 to 6 again with “RobberyRt” as your dependent variable.

CHAPTER 12 SPSS Exercises ➤ Explore additional data sets

<i>Data for Exercise</i>	
<i>Data Set</i>	<i>Description</i>
2013 YRBS.sav	The 2013 YRBS, short for Youth Risk Behavior Survey, is a national study of high school students. It focuses on gauging various behaviors and experiences of the adolescent population, including substance use and some victimization.
<i>Variables for Exercise</i>	
<i>Variable Name</i>	<i>Description</i>
WeaponCarrying	A scale based on three questions asking how often the student carries a weapon in and outside of school. High values indicate carrying weapons more frequently.
Qn24	A binary indicator for if the respondent had been bullied during the past year where 1 = yes and 0 = no.
Qn25	A binary indicator for if the respondent had been bullied online during the past year where 1 = yes and 0 = no.

Age	A respondent's age in years, ranging from 12 to 18+.
Sex	A binary variable where 1 = female and 2 = male.
White	A binary variable where 0 = non-White and 1 = White.
Grade	The grade that the respondent was in, ranging from 9th grade to 12th grade.
HostileEnvironment	An index assessing a student's experience of hostility in his or her school. This includes feeling unsafe in school and getting in any fights or being threatened. Higher values indicate higher levels of hostility experienced.

1. Why do some students carry weapons inside and outside of school? Who are those students? Dr. Smartypants is at it again: He is wondering if perhaps students who are bullied are the ones who are most prone to carrying weapons. To study this, he asks you to run a regression model with two indicators of if the student had been bullied online or offline. He hypothesizes that offline bullying is more likely to cause weapon carrying since it is in person. In his theory, offline bullying is unlikely to have an effect:
 - a. What are your null and research hypotheses? Be sure to include hypotheses for both independent variables and your regression model F test.
 - b. **Multiple Regression in SPSS:** As with single-variable regression, begin by selecting analyze->regression->linear. Place the variable "WeaponCarrying" in the dependent variable spot and both "Qn24" and "Qn25" in the independent variable slot. That's it! Select ok. You'll notice that the output is all but identical to the output provided within this chapter's text!
 - c. Use this output to answer the following:
 - i. What does the model's ANOVA table tell you in this case?
 - ii. What proportion of the variance in the weapon carrying scale is explained by these two variables?
 - iii. What do the slope coefficients for each tell us? Recall that these are binary indicators, so a 1-unit change in this case reflects the difference between not bullied students and bullied students. The hypothesis tests?
 - iv. How do you interpret the intercept in this case?
 - v. Is Dr. Smartypants's theory supported by these results?
2. Dr. Smartypants says he'd like to control for some additional variables to minimize the risk of confounding. In particular, he wants to control for student age, being male, year in school, and if the student is White versus non-White. To do this, include the variables age, sex, White, and grade in the model:
 - a. Does this change your results or conclusions?
 - b. How has your model R -squared changed?
 - c. Do you find any interesting or new effects in the independent variables you have added?
 - d. Take a look at the column called "standardized coefficients." These are your beta weights. What coefficients have the strongest effect on the dependent variable?
 - e. Briefly summarize your results for Dr. Smartypants. Are there any confounders or concerns that need to be considered?
3. You decide to check if multicollinearity may be an issue for your analysis. One simple way to look at this is to see if your independent variables are strongly correlated with one another. Consider a correlation coefficient of .5 to be evidence of multicollinearity in this case:
 - a. Do you find any evidence of multicollinearity? If so, explain why those two variables might be collinear.
 - b. If multicollinearity was found, reestimate your multiple regression model with one of the problem variables removed. Have your substantive conclusions changed?
4. You realize there is a potential confounder at play here. Maybe instead of bullying being the "cause" of weapon use, it's actually a sense of fear and hostility in the school. If a person gets bullied, they might get in fights or just feel

unsafe. Perhaps it is students who are exposed to a great deal of conflict and hostility who are carrying weapons. Let's make sure the bullying effects are still present after controlling for that confounder!

- a. Estimate your multiple regression model with the bullying variables age, sex, and White. This time, also add the variable "HostileEnvironment."
- b. Have your results (not) changed? Does this model fit better than earlier models? What is your strongest effect in the model based on your standardized coefficients and t statistics?
- c. Write a memo to Dr. Smartypants that uses the results of this model to either support or critique his theory presented in question 1. Be sure to include your model R -squared and how it has or hasn't changed, what happens when you control for the confounder, and the "why" that explains why the variables are related to the dependent variable.
- d. If you were tasked with developing an intervention to prevent weapon carrying in school, what would you recommend? What should be prioritized?
- e. Have you met the criteria for causality in this final model? If no, what more must be done?