Figure 1.6 Variable View Window for GSS2010

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	abany	Numeric	1	0	ABORTION IF...	{0, IAP}...	0, 8, 9	8	Right	Nominal	Input
2	abdefect	Numeric	1	0	STRONG CHAN...	{0, IAP}...	0, 8, 9	8	Right	Nominal	Input
3	abhlth	Numeric	1	0	WOMANS HEAL...	{0, IAP}...	0, 8, 9	8	Right	Nominal	Input
4	abnomore	Numeric	1	0	MARRIED--WA...	{0, IAP}...	0, 8, 9	8	Right	Nominal	Input

SPSS PROBLEM

Based on the *Utilities-Variables* option, review the variables from the GSS10SSDS. Can you identify three nominal variables, three ordinal variables, and at least one interval-ratio variable? Based on the information in the dialog box or Variable View window, you should be able to identify the variable name, variable label, and category values. You can do the same for the other data sets: MTF11SSDS, HINTS12SSDS, and GLOBAL13SSDS.

CHAPTER EXERCISES

1. In your own words, explain the relationship of data (collecting and analyzing) to the research process. (Refer to Figure 1.1.)
2. Construct potential hypotheses or research questions to relate the variables in each of the following examples. Also, write a brief statement explaining why you believe there is a relationship between the variables as specified in your hypotheses.
   a. Gender and educational level
   b. Income and race
   c. The crime rate and the number of police in a city
   d. Life satisfaction and marital status
   e. A nation's military expenditures as a percentage of its gross domestic product and that nation's overall level of security
   f. Care of elderly parents and ethnicity
3. Determine the level of measurement for each of the following variables:
   a. The number of people in your family
   b. Place of residence classified as urban, suburban, or rural
   c. The percentage of university students who attended public high school
   d. The rating of the overall quality of a textbook, on a scale from "Excellent" to "Poor"
   e. The type of transportation a person takes to work (e.g., bus, walk, car)
   f. Your annual income
   g. The U.S. unemployment rate
   h. The presidential candidate that the respondent voted for in 2012
4. For each of the variables in Exercise 3 that you classified as interval ratio, identify whether it is discrete or continuous.

5. Why do you think men and women, on average, do not earn the same amount of money? Develop your own theory to explain the difference. Use three independent variables in your theory, with annual income as your dependent variable. Construct hypotheses to link each independent variable with your dependent variable.

6. For each of the following examples, indicate whether it involves the use of descriptive or inferential statistics. Justify your answer.
   a. The number of unemployed people in the United States
   b. Determining students' opinion about the quality of food at the cafeteria based on a sample of 100 students
   c. The national incidence of breast cancer among Asian women
   d. Conducting a study to determine the rating of the quality of a new smartphone, gathered from 1,000 new buyers
   e. The average GPA of various majors (e.g., sociology, psychology, English) at your university
   f. The change in the number of immigrants coming to the United States from Southeast Asian countries between 2005 and 2010

7. Identify three social problems or issues that can be investigated with statistics. (One example of a social problem is criminal acts, such as murder.) Which one of the three issues would be the most difficult to study? Which would be the easiest? Why?

8. Construct measures of political participation at the nominal, ordinal, and interval-ratio levels. (*Hint:* You can use behaviors such as voting frequency or political party membership.) Discuss the advantages and disadvantages of each.

9. Variables can be measured according to more than one level of measurement. For the following variables, identify at least two levels of measurement. Is one level of measurement better than another? Explain.
   a. Individual age
   b. Annual income
   c. Religiosity
   d. Student performance
   e. Social class
   f. Attitude toward gun control

SPSS PROBLEMS

[GSS10SSDS and GLOBAL13SSDS]

1. Use the SPSS Frequencies command to produce a frequency table for the variable MARITAL as measured in the GSS10SSDS. How would you describe where most students in the sample were raised?
   a. What percentage of the sample is divorced?
   b. What percentage of the sample is married?
   c. What percentage of the sample would you describe as being currently single? (Include all relevant categories.)

2. The GSS2010 SSDS included a series of questions on respondent's attitudes about immigrants. In the chapter, we examined the relationship between race and attitudes about immigrants and jobs (IMMJOBS). The other two GSS variables include IMMCRIME and IMMAMECO.
   a. Run frequencies for all the three variables (including IMMJOBS).
   b. Prepare a general statement summarizing your results from the three frequency tables. Identify the level of measurement for each variable. How would you describe respondents' attitudes about immigrants?

3. Based on GSS10SSDS, produce the frequency table for the RACIDIMP, the importance of one's racial identity.
   a. What is the level of measurement for this variable?
   b. Identify two independent variables (included in the GSS10SSDS data set) that may be related to RACIDIMP. Explain the relationship between these variables and RACINIMP.

4. The GSS2010 SSDS asked respondents to report in their highest year of school (EDUC). Run the frequency table for this variable. Collapse this interval ratio variable into an ordinal measure (omitting those who did not respond to the question). How many categories do you have? Prepare a frequency and cumulative percentage table of your recoded EDUC variable.

5. Collapse the variables LABORRATEFEMALE and LABORRATEMALE (included in GLOBAL13SSDS) into ordinal measures. How many categories do you have? Prepare a frequency and cumulative percentage table of your recoded variables. What can you conclude about the difference in labor force participation between males and females?

CHAPTER EXERCISES

1. Suppose you have surveyed 30 people and asked them whether they are white (W) or nonwhite (N), and how many traumas (serious accidents, rapes, or crimes) they have experienced in the past year. You also asked them to tell you whether they perceive themselves as being in the upper, middle, working, or lower class. Your survey resulted in the raw data presented in the table below:
   a. What level of measurement is the variable race? Class?
   b. Construct raw frequency tables for race and for class.
   c. What proportion of the 30 individuals is nonwhite? What percentage is white?
   d. What proportion of the 30 individuals identified themselves as middle class?

Exercises

Race	Class	Trauma	Race	Class	Trauma
W	L	1	W	W	0
W	M	0	W	M	2
W	M	1	W	W	1
N	M	1	W	W	1
N	L	2	N	W	0
W	W	0	N	M	2
N	W	0	W	M	1
W	M	0	W	M	0
W	M	1	N	W	1
N	W	1	W	W	0
N	W	2	W	W	0
N	M	0	N	M	0
N	L	0	N	W	0
W	U	0	N	W	1
W	W	1	W	W	0

Source: Data based on GSS files for 1987 to 1991.

Notes: Race: W, white; N, nonwhite; Class: L, lower class; M, middle class; U, upper class; W, working class.

2. Using the data and your raw frequency tables from Exercise 1, construct a frequency distribution for class.
   a. Which is the smallest perceived class?
   b. Which two classes include the largest percentages of people?
3. Using the data from Exercise 1, construct a frequency distribution for trauma.
   a. What level of measurement is used for the trauma variable?
   b. Are people more likely to have experienced no traumas or only one trauma in the past year?
   c. What proportion has experienced one or more traumas in the past year?

5. A question on whether immigrants were good for America was included in the GSS 2010. Results are provided in the table below, noting the percentage who agree or strongly agree by political party (not all responses are reported here so totals will not add up to 100%). Do these data support the statement that people's views on immigration are related to their political party affiliation? Why or why not?

	Strong Democrat	*Independent*	*Strong Republican*
	%	%	%
Agree Strongly	9.1	7.7	2.1
Agree	47.7	34.6	47.9

6. How many hours per week do you spend on e-mail? In 2010, the GSS included a question on number of hours spent on e-mail. Data are presented here for a sample of 99 men and women.

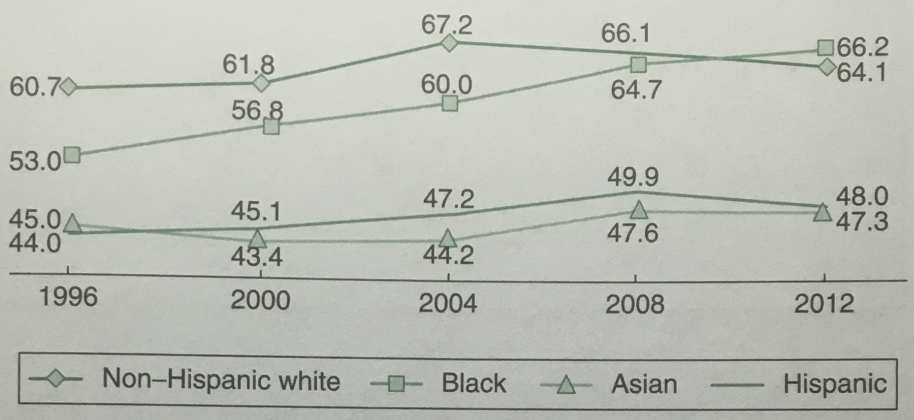
E-mail hours per week	*Frequency*
0	19
1	20
2	13
3	5
4	2
5	6
6	5
7	2
8	3
9	1
10 or more	23

   a. Compute the cumulative frequency and cumulative percentage distribution for the data.
   b. What proportion of the sample spent 3 hours or less per week on e-mail?
   c. What proportion of the sample spent 6 or more hours per week on e-mail?

7. The tables below present the frequency distributions for education by gender and race based on the GSS 2010. Use them to answer the following questions.

black voting rates exceeded the rates for non-Hispanic whites. Overall votes cast were higher in 2012 than 2008 (131,948,000 vs. 131,144,000—data not reported in the figure), an increase attributed to minority voters. Describe the variation in voting rates for the four racial and Hispanic origin groups.

Figure 3.23 Voting Rates by Presidential Election



Source: Thom File, *The Diversifying Electorate—Voting Rates by Race and Hispanic Origin in 2012 (and other recent elections),* Current Population Survey (P20-568), 2013, Figure 1.

2. We selected a sample of people from the International Social Survey Program (ISSP) 2000. Raw data are presented for their sex (SEX), social class (CLASS), and number of household members (HOMPOP). CLASS is a subjective measure, with respondents indicating L = *lower*, W = *working*, M = *middle*, and U = *upper.*

Sex	*Hompop*	*Class*	*Sex*	*Hompop*	*Class*
F	1	L	F	2	U
F	3	W	M	2	M
F	1	M	F	4	W
M	2	M	M	2	U
F	3	M	M	4	M
M	2	U	M	4	W
F	7	L	M	2	M

(Continued)

(Continued)

Sex	Hompop	Class	Sex	Hompop	Class
M	3	M	F	1	W
M	4	M	M	2	U
F	1	M	M	4	W
F	2	M	F	7	M
M	5	L	M	3	M
M	4	M	M	4	M
M	4	L	F	4	M
M	3	W	F	5	L

a. Construct a pie chart depicting the percentage distribution of sex. (*Hint:* Remember to include a title, percentages, and appropriate labels.)
b. Construct a pie chart showing the percentage distribution of social class.
c. Construct a graph with two pie charts comparing the percentage distribution of social class membership by sex.

3. We continue our analysis of the U.S. elderly population, examining household income for 2010. A histogram is presented in Figure 3.24. Write a brief statement describing the data.

4. Using the data from Exercise 2, construct bar graphs showing percentage distributions for sex and class. Remember to include appropriate titles, percentages, and labels.

5. Suppose you want to compare the number of household members for women and men (based on the ISSP data in Exercise 2).
   a. Construct a grouped bar graph (similar to Figure 3.4) to show the percentage distribution of the number of household members by sex.
   b. Which group reported the largest family size?
   c. Why shouldn't you construct a grouped bar chart showing the frequencies rather than the percentages?

6. Policy analysts have noted that the number of those without health insurance is increasing in the United States. Access to health insurance has been identified as an important social issue. Data from the National Center for Health Statistics (2013) are presented below (see Table 3.2), measuring the percent of persons with no health insurance for at least part of 2011 by selected characteristics. Note: Characteristic totals will not equal 100%.
   a. What can be said about who did not have health insurance in 2011? How does the percentage of those without health insurance vary by each demographic characteristic?
   b. For each variable, what would be the best way to graphically present the data?

option. Click *OK*. SPSS will create a great deal of output; all you need to do is find the appropriate frequency tables and measures of central tendency. For example, to find and report the frequency table for males who are pretty happy, look for the section with values of Happiness of Marriage = Pretty Happy and Respondent's Sex = Male.

b. Do you notice any gaps in hours spent watching TV between men and women at different levels of marital satisfaction?

CHAPTER EXERCISES

1. The following frequency distribution contains information about people's self-evaluations of their lives.

Respondent Assessment of Life	*Frequency*	*Percentage*	*Cumulative Percentage*
Exciting	470	48.3	48.3
Routine	444	45.6	93.9
Dull	60	6.1	100.0
Total	974	100.0	

Source: GSS, 2010.

   a. Find the mode.
   b. Find the median.
   c. Interpret the mode and the median.
   d. Why would you not want to report the mean for this variable?

2. Same-sex unions have increasingly become a heated political issue. The 2010 GSS asked respondents' opinions on homosexual relations. Four response categories ranged from "Always Wrong" to "Not Wrong at All." See the following frequency distribution:

Homosexual Relations	*Frequency*	*Percentage*	*Cumulative Percentage*
Always wrong	467	50.2	50.2
Almost always wrong	41	4.4	54.6
Sometimes wrong	76	8.2	62.8
Not wrong at all	346	37.2	100.0
Total	930	100.0	

(Continued)

State	2005	2009
Illinois	1,674	1,806
Minnesota	691	767
New Hampshire	185	217
New York	2,758	2,937
Washington	807	938

Source: U.S. Census Bureau, *The 2012 Statistical Abstract*, Table 147.

a. Calculate the mean number of Americans on Medicare in these eight states for both 2005 and 2009. How would you characterize the difference in the number of Americans on Medicare between 2005 and 2009? Does the mean adequately represent the central tendency of the distribution of Americans on Medicare in each year for these eight states? Why or why not?
b. Recalculate the mean for each year after removing Florida, Illinois, and New York from the table. Is the mean now a better representation of central tendency for the remaining five states? Explain.

7. U.S. households have become smaller over the years. The following table from the 2010 GSS contains information on the number of people currently aged 18 years or older living in a respondent's household. Calculate the mean number of people living in a U.S. household in 2010.

Household Size	Frequency
1	381
2	526
3	227
4	200
5	96
6	42
7	19
8	5
9	2
10	2
Total	1,500

8. In Exercise 6, you calculated the mean number of Americans on Medicare. We now want to test whether the distribution of Americans on Medicare is symmetrical or skewed.
a. Calculate the median and mode for each year, using all eight states. Based on these results and the means, how would you characterize the distribution of Americans on Medicare for each year?

b. Does the mean or median best represent the central tendency of each distribution? Why?
c. If you found the distributions to be skewed, what might be the statistical cause?

9. In Exercise 7, you examined U.S. household size in 2010. Using these data, construct a histogram to represent the distribution of household size.
   a. From the appearance of the histogram, would you say the distribution is positively or negatively skewed? Why?
   b. Now calculate the median for the distribution and compare this value with the value of the mean from Exercise 7. Do these numbers provide further evidence to support your decision about how the distribution is skewed? Why do you think the distribution of household size is asymmetrical?

10. Exercise 3 used GSS data on the number of hours worked per week for a sample of 32 Latino adults.
    a. Calculate the mean number of hours worked per week.
    b. Compare the value of the mean with those you have already calculated for the median. Without constructing a histogram, describe whether and how the distribution of weeks worked per year is skewed.

11. You listen to a debate between two politicians discussing the economic health of the United States. One politician says that the average income of all workers in the United States is $72,235; the other says that American workers make, on average, only $52,029, so Americans are not as well off as the first politician claims. Is it possible for both these politicians to be correct? If so, explain how.

12. Discuss the advantages and disadvantages of all three measures of central tendency. Are you confident that one of these three is the best measure of central tendency? If so, why?

13. Do male murder rates vary with country population? Investigate this question using the following data for selected countries grouped by population size, the top 10 countries and the bottom 10. Calculate the mean and median for each group of countries. Where is the murder rate for males the highest? Do the mean and median have the same pattern for the two groups?

2008–2010 Male Murder Rate per 100,000			
Top 10 by Population	*Murder Rate*	*Bottom 10 by Population*	*Murder Rate*
China	2.2	Vietnam	2.6
India	3.9	Egypt	2.2
United States	6.6	Germany	0.9
Indonesia	13.9	Turkey	8.6
Brazil	54.7	Iran	2.3
Pakistan	4.3	Democratic Republic of the Congo	35.8
Nigeria	18.2	France	1.9
Russia	29.1	United Kingdom	1.7
Japan	0.4	Italy	1.6
Mexico	23.0	South Korea	2.2

Source: United Nations Office on Drugs and Crime, *2011 Annual Report.*

a. What is the range of convictions in 1990? In 2009? Which is greater?
b. What is the mean number of convictions in 1990 and 2009?
c. Calculate the standard deviation for 1990 and 2009.
d. Which year appears to have more variability in number of convictions as measured by the standard deviation? Are the results consistent with what you found using the range?

4. Your task is to construct a report regarding criminal offenses investigated by U.S. attorneys by offense and year using the following data from the U.S. Department of Justice. Your report should include the appropriate measures of central tendency, measures of variability, and a few sentences comparing the number of criminal offenses in 2005 with the number of criminal offenses in 2009. Also, include an explanation for any reported difference between 2005 and 2009.

2005		*2009*	
Type of Offense	*No. of Suspects*	*Type of Offense*	*No. of Suspects*
Violent	5,485	Violent	5,463
Property	25,570	Property	26,161
Drug	40,038	Drug	37,721
Public order	21,583	Public order	23,067
Weapon	13,689	Weapon	11,749
Immigration	36,559	Immigration	88,313

Source: U.S. Department of Justice, *Federal Justice Statistics 2009,* Table 4.

5. The output below depicts data for projected elderly population change in Midwestern and Western states between 2008 and 2015 from Table 5.4.

Descriptives

	Region		Statistic
Population_Change	Midwest	Mean	13.600
		Std. Deviation	2.7831
		Minimum	9.4
		Maximum	19.2
		Range	9.8
		Interquartile Range	3.7
	West	Mean	28.277
		Std. Deviation	10.6948
		Minimum	13.8
		Maximum	50.0
		Range	36.2
		Interquartile Range	17.3

Exercises

a. Compare the range for the western states to that of the Midwest. Which region had a greater range?
b. Examine the IQR for each region. Which is greater?
c. Use the statistics to characterize the variability in population increase of the elderly in the two regions. Does one region have more variability than another? If yes, why do you think that is?

6. Occupational prestige is a statistic developed by sociologists to measure the status of one's occupation. Occupational prestige is also a component of what sociologists call socioeconomic status, a composite measure of one's status in society. On average, people with more education tend to have higher occupational prestige than people with less education. We investigate this using the 2010 GSS variable PRESTG80 and the Explore procedure to generate the selected SPSS output shown in Figure 5.15.

Figure 5.15 Descriptive Statistics for Occupational Prestige Score by Highest Degree Earned

PRESTG80			Statistic
RS OCCUPAIONAL PRESIGE SCORE (1980)	High School Diploma	Mean	40.59
		Median	40.00
		Std. Deviation	11.419
		Minimum	17
		Maximum	75
		Range	58
		Interquartile Range	17
	Bachelor's Degree	Mean	50.95
		Median	51.00
		Std. Deviation	12.930
		Minimum	23
		Maximum	75
		Range	52
		Interquartile Range	23

a. Note that SPSS supplies the IQR, the median, and the minimum and maximum values of each group. Looking at the values of the mean and median, do you think the distribution of prestige is skewed for respondents with a high school diploma? For respondents with a bachelor's degree? Why or why not?
b. Explain why you think there is more variability of prestige for either group, or why the variability of prestige is similar for the two groups.

7. The U.S. Census Bureau collects information about divorce rates. The following table summarizes the divorce rate for 10 U.S. states in 2007. Use the table to answer the questions that follow.

State	*Divorce Rate per 1,000 Population*
Alaska	4.3
Florida	4.7
Idaho	4.9

(Continued)

(Continued)

State	Divorce Rate per 1,000 Population
Maine	4.5
Maryland	3.1
Nevada	6.5
New Jersey	3.0
Texas	3.3
Vermont	3.8
Wisconsin	2.9

Source: U.S. Census Bureau, *Statistical Abstract of the United States: 2010*, Table 126.

a. Calculate and interpret the range and the IQR. Which is a better measure of variability? Why?
b. Calculate and interpret the mean and standard deviation.
c. Identify two possible explanations for the variation in divorce rates across the 10 states.

8. The respondents of the 2007 HINTS reported their psychological distress on a scale between 0 and 24. In the table below, you will see separate data on two groups of respondents' distress scores: those who have ever been diagnosed as having cancer, and those who have not.

	Psychological Distress Score	
	Diagnosed	Not Diagnosed
$\bar{Y}$	3.9	4.87
ΣY	729	5,849
$\Sigma(Y-\bar{Y})^2$	3,059.14	25,180.20
N	187	1,200

a. Calculate the variance and standard deviation from these statistics for both groups.
b. What can you say about the variability in the distress scores for those respondents who have been diagnosed as having cancer and those who have not? Why might there be a difference? Why might there be more variability for one group than for the other?
c. Was it necessary in this problem to provide you with the mean value to calculate the variance and standard deviation?

9. You are interested in studying the variability of crimes committed (including violent and property crimes) and police expenditures in the eastern and Midwestern United States. The U.S. Census Bureau collected the following statistics on these two variables for 21 states in the East and Midwest in 2008.

Exercises