

## Chapter 2 Exercises

Name \_\_\_\_\_

Date \_\_\_\_\_

1. Create a new data set by polling 5 other people and asking the following questions:

RESPOND. What are the last four digits of your phone number?

EXER1. Do you exercise regularly? (Yes or No)

EXER2. How many hours a week would you say you exercise?

EXER3. Do you participate in team sports? (Yes or No)

EXER4. What is your favorite sport to play?

Create a new data set using these data and incorporating the following labels.

Variable Name	Type	Label	Value Labels	Measure
RESPOND	String20	Respondent Name	None	Nominal
EXER1	Numeric	Exercise Regularly?	0=no 1=yes	Nominal
EXER2	Numeric	Hours/Week of Exercise	None	Scale
EXER3	Numeric	Participate in Team Sports?	0=no 1=yes	Nominal
EXER4	String 20	Favorite Sport	None	Nominal

Enter the responses and save the data as “EXERCISE.SAV”. Be sure to save it to a location that you have access to. Show that you have set up the file with the following command and follow your instructor’s request, such as print the output:

```
File
  Display Data File Information
  Working File
```

**Variable Information**

Variable	Position	Label	Measurement Level	Role	Column Width	Alignment	Print Format	Write Format
RESPOND	1	Respondent Name	Nominal	Input	8	Left	A20	A20
PHONE	2	Phone Number	Nominal	Input	8	Right	F8.2	F8.2
EXER1	3	Exercise Regularly?	Nominal	Input	8	Right	F8.2	F8.2
EXER2	4	Hours/Week of Exercise	Scale	Input	8	Right	F8.2	F8.2
EXER3	5	Participate in Team Sports	Nominal	Input	8	Right	F8.2	F8.2
EXER4	6	Favorite Sport	Nominal	Input	8	Left	A20	A20

Variables in the working file

**Variable Values**

Value	Label
EXER1 .00	no
1.00	yes
EXER3 .00	no
1.00	yes

2. Open the data STATES10.SAV. Examine the data set in the *Variable View Window* and locate some variables that measure the following concepts. Recall that the variable names begin with a 2-letter code for the general topic of the variable. (Double click within the box of variables and scroll up or down to view all of the variables.)

A. Economic Prosperity

*ECS96 Per Capita Gross Domestic Product: 2008*  
*ECS100 Per Capita Personal Income: 2008*  
*ECS103 Median Household Income: 2008*  
*ECS103 Median Household Income: 2008*  
*ECS105 Personal Bankruptcy Rate per 100,000: 2009*  
*ECS445 Homeownership Rate: 2008*

B. Substance Abuse

*HTC254 Percent of High School Students Who Drink Alcohol: 2007*  
*HTC255 Percent of High School Students Who Use Marijuana: 2007*  
*HTH180 Death Rate per 100,000 by Alcohol-Induced Deaths: 2006*  
*HTH503 Adult Per Capita Beer Consumption: 2007*  
*HTH505 Adult Per Capita Wine Consumption: 2007*  
*HTH507 Adult Per Capita Distilled Spirits Consumption: 2007*  
*HTH508 Percent of Adults Who Do Not Drink Alcohol: 2008*  
*HTH510 Percent of Adults Who Smoke: 2008*  
*HTH515 Percent of Population Who are Illicit Drug Users: 2007*  
*HTS420 Adult Per Capita Alcohol Consumption: 2007*

C. Educational Attainment

*EDS127 Percent of Public School 4th Graders Proficient or Better in Reading: 2007*  
*EDS128 Percent of Public School 8th Graders Proficient or Better in Reading: 2007*  
*EDS129 Percent of Public School 4th Graders Proficient or Better in Math: 2009*  
*EDS130 Percent of Public School 8th Graders Proficient or Better in Math: 2009*  
*EDS131 Percent of Population Graduated from High School: 2008*  
*EDS134 Public High School Drop Out Rate: 2007*  
*EDS135 ACT Average Composite Score: 2009*  
*EDS151 Enrollment Rate per 1,000 Aged 18-24 in Institutions of Higher Education: 2006*  
*EDS154 Percent of Population With a Bachelors Degree or More: 2008*

3. Your state’s Department of Education hires you as a statistical consultant to help design a data set. They have already sent a survey to school superintendents, and need you to design their data set so they can begin entering information. Below is an example of a survey returned from one school district. Using the survey, answer the questions below, showing how you would design the data set.

1. Number of elementary schools:	8
2. Number of middle schools:	4
3. Number of high schools:	2
4. Average spending per pupil:	\$7500
5. Student/Teacher ratio:	19/1
6. High School drop-out rate:	10%
7. Does the High School offer AP classes?	<u>Yes</u>
8. Type of district: (circle one)	Rural <u>Suburban</u> Urban

In this data set, what will be a case? School district or Superintendent

Design variable names, labels, and measurement types for each of the variables in the study.

Variable Name	Variable Label	Value Labels	Measurement Type
---------------	----------------	--------------	------------------

- 1. *Answers will vary.*
- All variables should have names 8 digits or fewer.*
- All variables should have variable labels.*
- Variables 7 and 8 should have value labels.*
- Variables 1,2,3,4,5,and 6 are scale variables.*
- Variables 7 and 8 are categorical variables,*

4. Open the GSS08 data and examine the variable SATFIN (Satisfaction with Financial Situation). Can you identify the values and the value labels?

Value	<u>1</u>	Value label	<u>Satisfied</u>
Value	<u>2</u>	Value label	<u>More or Less</u>
Value	<u>3</u>	Value label	<u>Not at All Sat</u>
Value	<u>8</u>	Value label	<u>DK</u>
Value	<u>9</u>	Value label	<u>NA</u>

Note, in the GSS08 Data, IAP means “Inapplicable,” DK means “Don’t Know” and NA means “No Answer”.

5. A novice researcher begins an analysis of the GSS08 data and wants to understand the distribution of income in the United States. She looks at the data in the *Data View* and sees values that range mostly from 1 to 25, but occasionally also sees the number 98 (you can do this by finding the variable INCOME06 in the GSS08 data). She explains to her instructor that the data must be wrong because almost no one has an income of “10.” Her instructor tells her the data are fine, and asks her to come back when she can tell him how much a family income is for an observation that is listed as “10”.

What does “10” in the INCOME06 variable indicate? \$12,500-\$14,999

What is the reason for the novice researcher’s confusion?

*The confusion stems from thinking of the data value as being equivalent to the dollar income. In this case, the 10 indicates a value assigned to an income range. In the GSS08 data, “10” is the value assigned to identify individuals with incomes that range from \$12,500-\$14,999 in the variable INCOME98.*

6. Open the file FAMILY.SAV that you saved earlier and design 4 new variables to put into the data set and enter in the values. Try to vary the structure of the variables so that you include at least one scale variable, one ordinal variable, and one nominal variable. For example (don't use this one!), we could add the variable "FISH" with the label "Likes to Fish", the value labels "0=NO, 1=YES", and the measurement type "Nominal". Enter the responses and save the data as "FAMILY2.SAV". Show that you have set up the file with the following command and follow your instructor's directions, such as print the output:

*File*  
*Display Data File Information*  
*Working File*

*Answers will vary.*