Chapter 7

Basic Ideas of Probability

The notion of chance is commonly used when defining probability. In statistical analysis, a random sample of individuals from the population is chosen in such a way that all possible sample sets have the same chance, or same probability, of being chosen. Although SPSS for Windows is designed primarily to be used for data analysis and not for evaluating probability functions per se, it is possible to demonstrate certain probability concepts with the program. This chapter illustrates "tossing a coin" and "rolling a die" using SPSS. All of the procedures discussed in this chapter involve *sampling with replacement;* using this approach, the probability of a particular outcome is not changed by of the outcome(s) that precede it.

7.1 PROBABILITY IN TERMS OF EQUALLY LIKELY CASES

This section demonstrates how to simulate equally likely outcomes of simple operations such as tossing a coin or dealing a card from a shuffled deck. The Bernoulli distribution is the distribution of a variable that can take one of two values — 0 or 1. There are several Bernoulli distributions, differing with respect to the probability associated with the values. If the probability is .5 that a random draw from this distribution will be a 1 (and .5 that it will be a 0), we can say that sampling from this distribution is the same as tossing a fair coin. For

our purposes, obtaining a value of 1 corresponds to tossing a "head," and a value of 0 corresponds to a "tail."

To simulate coin tossing on SPSS:

1. Click on **File** from the menu bar.
2. Click on **New** from the pull-down menu.
3. Click on **Data** from the pull-down menu.
4. You need to compute a variable that has a value of 1 or 0 (a Bernoulli variable). SPSS will not permit you to compute any variable without having an active data set, however. To create such a data set, you must type some number (e.g., 999) in the first cell of the first column of the data file.
5. Click on **Transform** from the menu bar.
6. Click on **Compute** from the pull-down menu to open the Compute Variable dialog box (see Fig. 7.1).
7. Type in the name of the new variable (e.g., "coin") in the Target Variable box.
8. Click on **Random Numbers** in the Function group box and then double-click on **Rv.Bernoulli** in the Functions and Special Variables box. Notice that the function now appears in the Numeric Expression box.
9. Select a specific Bernoulli distribution by indicating the probability of obtaining a value of 1. Type **.5** where the question mark is.
10. Click on **OK** to run the procedure.

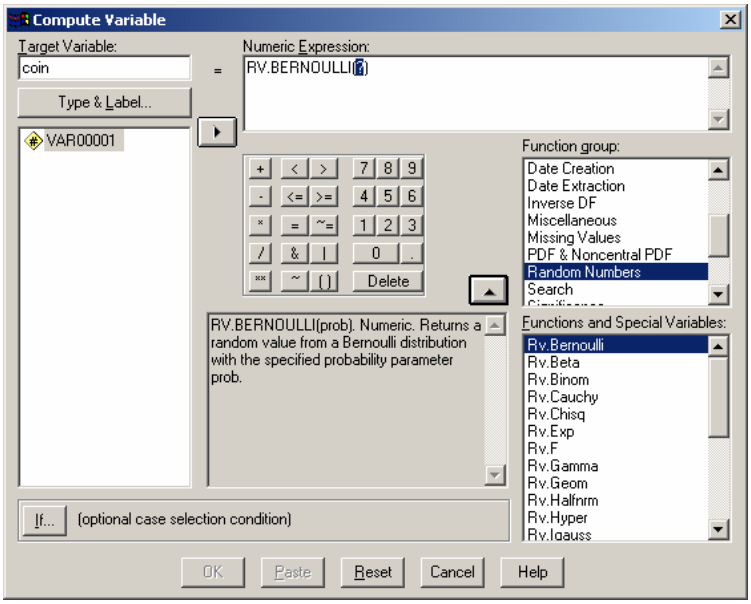

Figure 7.1 Compute Variable Dialog Box

Note that the value of the new variable, "coin," has a value of 1 (head) or 0 (tail). This represents the result of the coin toss.

You can also simulate rolling a die with a similar procedure. The random variable that corresponds to this is from a discrete uniform distribution, with values ranging from 1 to 6. SPSS has only the continuous uniform distribution as a random variable function. To make SPSS roll a die, follow steps 1–6 above, and then:

1. Type in the name of the new variable (e.g., "die") in the Target Variable box.
2. Click on **Arithmetic** in the Function group box and double-click on **Rnd** in the Functions and Special Variables box. This is the round function, which rounds continuous decimal numbers to integers.
3. Notice that "RND(?)" appears in the Numeric Expression box. Indicate the value or expression you wish to round. Click on Random Numbers in the Function group box and then double-click on **Rv.Uniform** in the Functions and Special Variables box. It should automatically replace the question mark from the "RND" expression.
4. Replace the question marks in the expression with **"1, 6"**.
5. Click on **OK** to run this procedure.

In the column labeled die in the data file, there will be an integer between 1 and 6, representing the roll of the die.

7.2 RANDOM SAMPLING; RANDOM NUMBERS

Drawing samples at random is often subject to biases (e.g., the deck was not shuffled adequately or the coin was slightly imperfect). By using numerical methods, the computer can simulate events with specific probabilities that are free from bias.

To generate several random numbers from a distribution, you can repeat the procedure in Section 7.1 several times. To do so, simply modify step (4) in Section 7.1 by entering a number for multiple cases for the variable. For instance, if you want to perform 10 coin tosses, you need to enter some number, all 999's for instance, in the first 10 cells of the first column. Then, compute the "coin" variable exactly as described above. You will obtain 10 values, each either 0 (tails) or 1 (heads), representing the result of each of ten coin tosses. Similarly, you may simulate 10 rolls of a die by using 1 and 6 as the minimum and maximum values, respectively, in the "RV.UNIFORM(min,max)" expression.

Chapter Exercises

7.1 Following the instructions in Section 7.2, simulate 20 coin tosses with SPSS.

- **a.** Did you obtain 10 heads (1's) and 10 tails (0's)? Why or why not?
- **b.** Would you have come closer to an equal split with 50 tosses? With 10 tosses? Why or why not?

7.2 Think of a number between 1 and 5. Direct SPSS to pick a number at random within the range 1 through 5. (Hint: use the uniform distribution).

- **a.** Did the two numbers coincide? What is the probability that they would be the same?
- **b.** Repeat this procedure 10 times. How many matches were there? How many did you expect to obtain (based on the law of probability)?