Chapter 8

Probability Distributions

This chapter demonstrates how to use SPSS for Windows to generate probability distributions. Probabilities of events can be described for categorical variables as shown in Chapter 7 (e.g., the likelihood that a coin toss will be a head), or numerically as described in this chapter (e.g., the likelihood that a person regularly eats 50 candy bars per year). We will focus on properties of the standard normal distribution, but SPSS can perform the same tasks for other probability distributions as well. Using SPSS, we demonstrate how to:

- find the probability associated with a given value of a standard normal variable;
- find the value of a standard normal variable associated with a given probability.

8.1 FAMILY OF STANDARD NORMAL DISTRIBUTIONS

Many numerical variables have distributions that look similar — a smooth curve in the shape of a bell. This shape is based on a probability distribution known as the normal distribution. The Central Limit Theorem tells us that whenever a variable is made up of many separate "components," each of which can have two or more values, the resulting variable will have approximately a normal probability distribution. The greater the number of components, the more perfectly normal the probability distribution for the variable will be. A normally shaped distribution with mean 0 and standard deviation 1 is called the standard normal distribution (see Chapter 4).

Finding Probability for a Given z-Value

When examining probability distributions, it is important to be able to determine the percentage of the distribution that lies within a certain interval. We know, for instance, that 50% of the area under the standard normal curve lies below the point 0. In other words, the probability that a random variable drawn from a standard normal distribution will be less than 0 is .50.

The cumulative distribution function in SPSS for Windows will compute these probabilities. The procedure for obtaining these computations is detailed below:

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. SPSS will not permit you to compute a new variable without having an active data set. So, "activate" the Data window by typing 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.
7. Type in the name of the new variable (e.g., "probilty") in the Target Variable box.
8. Click on **CDF & Noncentral CDF** in the Function group box and double-click on **Cdf.Normal** in the Functions and Special Variables box.
9. Notice the function in the Numeric Expression box. The first question mark (the "q" parameter) represents the point on the distribution for which you wish to obtain a cumulative probability estimate. For the first example, we will find the cumulative probability distribution for 0. Therefore, modify the expression to read "**CDF.NORMAL(0,0,1).**"
10. Click on **OK** to run the procedure.

SPSS will return the value of .5 as the value of the "probilty" variable.

Repeat this procedure, this time for the value 1. You can do so by starting at step (5) and changing the "q value" from 0 to 1. (Note: When you click on **OK** SPSS will prompt you to indicate whether or not you wish to "Change Existing Variable?") Then click on **OK** to run the procedure. The value of the "probilty" variable should change to .84, indicating that 84% of the distribution is below 1 standard deviation above the mean.

Finding a z-Value for a Given Probability

There may also be instances when you wish to determine the point on a probability distribution associated with a given cumulative probability. For instance, the z-value that has 50% of the standard normal distribution below it is 0, and a z-value that has 95% of the standard normal distribution below it is 1.64. The function in SPSS that will return the z-values for given probabilities is the inverse distribution function.

The procedure for obtaining this result is similar to that outlined previously. Follow steps 1–6 above, and then:

1. Type in the name of the new variable (e.g., "value") in the Target Variable box.
2. Click on **Inverse DF** in the Function group box and double-click on **Idf.Normal** in the Functions and Special Variables box.
3. To find the value on the standard normal distribution that has 95% of the values below it, modify the function in the Numeric Expression box by entering the value **.95** for the probability parameter, the value of **0** for the mean parameter, and the value of **1** for the stddev parameter.
4. Click on **OK** to run the procedure.

You should obtain the value of 1.64.

Chapter Exercises

8.1 Using SPSS, find the value on the standard normal distribution that has 2.3% of the distribution below it.

8.2 Sketch the curve of a normal distribution with mean 2 and standard deviation 1. How much of this distribution is below the value 2? Verify your answer using SPSS.

8.3 Estimate the value on the standard normal distribution that has 60% of the distribution below it. Use SPSS to evaluate your estimate.

8.4 Repeat Exercise 8.2 using a normal distribution with mean 0 and standard deviation 2.

8.5 Using SPSS, determine the proportion of the standard normal curve that is:

a. above 1.

b. between –2 and +2.

Hint: SPSS will not compute these areas directly; you must perform some simple computations by hand.