**APPENDIX B**

**GENERATING DISTRIBUTIONS FOR CHAPTERS 8 and 9 USING SPSS SYNTAX**

In this appendix, we describe the output that will be generated by executing each of four SPSS syntax programs used to generate figures in the text and give instructions for how to execute each of these programs.

1. ***IDGENER.SPS*** is a file that contains the syntax program to generate a new data set of ID values for 75 cases. If you would like a different number of ID’s to be generated, simply change the value of 75 to that other number. This program is useful also as the first step in generating the binomial distributions of chapter 8, since these all require there to be data values in the data editor before applying the binomial distribution function.
2. ***SAMPDIS.SPS*** is a file that contains the syntax program to generate a sampling distribution of means. In particular, this program generates the sampling distribution of means illustrated in Figure 9.2 that is based on the small population of ten values given in the text. In this syntax program, the ten population values are included within the program itself and are accessed by the DATA LIST, BEGIN DATA and END DATA commands.
3. ***NORMDIS9\_3.SPS*** is a file that contains the syntax program to generate the normal distribution illustrated in Figure 9.3.
4. ***NORMDIS9\_4.SPS*** is a file that contains the syntax program to generate the normal distribution illustrated in Figure 9.4.
5. ***POSSKEWDIS9\_5.SPS*** is a file that contains the syntax program to generate the positively skewed distribution illustrated in Figure 9.5.
6. ***SAMPDISVER2.SPS*** is file that contains another version (Version 2) of the syntax program *SAMPDIS.SPS*. Like the original version, Version 2 generates a sampling distribution of means, but unlike the original version, Version 2 gets the population values from which it generates the sampling distribution from an external data file (using the GET FILE command) located on the CD accompanying the text. If your CD drive is designated as a:, the file has address: ‘a:\data\positively skewed population.sav’. This external data file contains the 1,000 positively skewed values graphed in Figure 9.5. The histogram of the sampling distribution of means generated by this program is given in Figure 9.6.

**(1) CREATING A NEW DATA SET FILE WITH ID VALUES FOR 75 CASES**

Directions for use: Open the file labeled ***IDGENER.SPS*** in the SPSS syntax window from the CD that accompanies the text. Click **Run**, **All**. A new file containing 75 ID values will be generated.

The syntax file *IDGENER.SPS* appears below:

input program.

loop #i=1 to 75.

compute id = #i.

end case.

end loop.

end file.

end input program.

execute.

**(2) THE SPSS SYNTAX PROGRAM (CALLED, IN GENERAL, A MACRO) TO GENERATE THE SET OF 50,000 SAMPLE MEANS USED TO FORM THE SAMPLING DISTRIBUTION OF MEANS GRAPHED AS THE HISTOGRAM OF FIGURE 9.2**

In this program, called ***SAMPDIS.SPS***, the number of samples to be generated equals nsamp (which equals 50,000 in this version of the program); the size of the population is represented by the variable nsize (which equals 10 in this version of the program); each of the 50,000 samples generated is specified to be of size samsize (which equals 8 in this version of the program). Accordingly, in this version of the program, the sampling distribution of means is based on 50,000 samples, each of size N = 8, drawn randomly with replacement from the population: 0, 1, 3, 3, 5, 7, 7, 7, 8, 10 found in Chapter 9. These ten population values are input directly into the syntax program between the BEGIN DATA and END DATA commands.

Directions for use: Open the file labeled *SAMPDIS.SPS* in the SPSS syntax window from the CD that accompanies the text. Click **Run**, **All**. The 50,000 sample means that are generated will appear in the data editor and will also be assigned the file name bootmean.sav. You may then construct a histogram of these data.

If you wish to adjust the population, change the set of data values between “begin data” and “end data”. If you change the size of the population, make the appropriate adjustment to the value of nsize. If you wish to change the number of sample means generated, change the value nsamp. If you wish to change the size of each of the samples randomly selected, change the value of samsize.

The syntax file *SAMPDIS.SPS* appears below:

DEFINE bootmn (nsamp !tokens(1) / nsize !tokens(1) /samsize !tokens(1)

/ bootvar !tokens(1) / outfile !tokens(1)).

sort cases by !bootvar.

vector data (!nsize).

compute data ($casenum) = !bootvar.

compute nobreak = 1.

aggregate outfile \*

/break nobreak

/data1 to !concat(data,!nsize) = max(data1 to !concat(data,!nsize)).

vector data = data1 to !concat(data,!nsize).

vector tmp(!nsize).

loop #p = 1 to !nsamp.

loop #q = 1 to !nsize.

compute tmp(#q) = 0.

end loop.

loop #I = 1 to !samsize.

compute id = trunc(uniform(!nsize) + 1).

compute tmp(id) = tmp(id) + 1.

end loop.

compute mean = 0.

loop #j = 1 to !nsize.

compute mean = mean + data(#j)\*tmp(#j).

end loop.

compute mean = mean/!samsize.

\*save to an outfile\*.

xsave outfile !quote(!outfile)/keep mean.

end loop.

execute.

!ENDDEFINE.

data list /x 1-2.

begin data.

0

1

3

3

5

7

7

7

8

10

end data.

bootmn nsamp = 50000 nsize = 10 samsize = 8 bootvar = x outfile = bootmean.

get file "bootmean".

execute.

**(3) THE SPSS SYNTAX PROGRAM TO GENERATE THE SET OF 1,000 NORMALLY DISTRIBUTED SCORES WITH MEAN = 15 AND SD = 3 AS ILLUSTRATED BY THE HISTOGRAM OF FIGURE 9.3**

In this program, called ***NORMDIS9\_3.SPS***, 1,000 probabilities are generated to be equally spaced on the interval from 0 to 1. The spacing is controlled by the statement compute prob = prob + 1/1000. These probabilities are then input into the idf.normal function to produce a set of 1,000 normally distributed scores with mean and sd values specified to be 15 and 3, respectively.

Directions for use: Open the file labeled *NORMDIS9\_3.SPS* in the SPSS syntax window from the CD that accompanies the text. Click **Run**, **All**. The set of 1,000 normally distributed scores with mean = 15 and sd = 3 will appear in the data view window of the SPSS data editor. You may then construct a histogram of these data in the usual way using Graphs/Histogram, placing SCORE in the Variable box, to obtain the histogram of Figure 9.3.

The syntax file *NORMDIS9\_3.SPS* appears below:

new file.

input program.

compute prob = -(1/2000).

loop k = 1 to 1000.

compute mean = 15.

compute sd = 3.

compute prob = prob + 1/1000.

compute score = idf.normal(prob,mean,sd).

leave k, prob, score.

end case.

end loop.

end file.

end input program.

execute.

**(4) THE SPSS SYNTAX PROGRAM TO GENERATE THE SET OF 1,000 NORMALLY DISTRIBUTED SCORES WITH MEAN = 15 AND SD = 3 AS ILLUSTRATED BY THE HISTOGRAM OF FIGURE 9.4**

In this program, called ***NORMDIS9\_4.SPS***, 1,000 probabilities are generated to be equally spaced on the interval from 0 to 1. The spacing is controlled by the statement compute prob = prob + 1/1000. These probabilities are then input into the idf.normal function to produce a set of 1,000 normally distributed scores with mean and sd values specified to be 15 and 0.75, respectively.

Directions for use: Open the file labeled *NORMDIS9\_4.SPS* in the SPSS syntax window from the CD that accompanies the text. Click **Run**, **All**. The set of 1,000 normally distributed scores with mean = 15 and sd = 0.75 will appear in the data view window of the SPSS data editor. You may then construct a histogram of these data in the usual way using Graphs/Histogram, placing score in the Variable box, to obtain the histogram of Figure 9.4.

The syntax file *NORMDIS9\_4.SPS* appears below:

new file.

input program.

compute prob = -(1/2000).

loop k = 1 to 1000.

compute mean = 15.

compute sd = 0.75.

compute prob = prob + 1/1000.

compute score = idf.normal(prob,mean,sd).

leave k, prob, score.

end case.

end loop.

end file.

end input program.

execute.

**(5) THE SPSS SYNTAX PROGRAM TO GENERATE THE SET OF 1,000 POSITIVELY-SKEWED DISTRIBUTED SCORES WITH MEAN = 8 AND SD = 4 AS ILLUSTRATED BY THE HISTOGRAM OF FIGURE 9.5**

In this program, called ***POSSKEWDIS9\_5.SPS***, 1,000 probabilities are generated to be equally spaced on the interval from 0 to 1. The spacing is controlled by the statement compute prob = prob + 1/1000. These probabilities are then input into the idf.chisq function to produce a set of 1,000 positively-skewed scores with mean specified to be 8.

If the mean of this distribution (called a chi-square distribution) equals 8, the sd will automatically be equal to 4 because the sd is a function of the mean (sqrt(2\*mean) in this distribution. Accordingly, unlike for the normal distribution, which requires specification of both the mean and standard deviation, only the mean needs to be specified in the chi-square distribution. In this example, sqrt(2\*mean) = sqrt(2\*8) = sqrt(16) = 4.

Directions for use: Open the file labeled *POSSKEWDIS9\_5.SPS* in the SPSS syntax window from the CD that accompanies the text. Click **Run**, **All**. The set of 1,000 normally distributed scores with mean = 8 and sd = 4 will appear in the data view window of the SPSS data editor. You may then construct a histogram of these data in the usual way using Graphs/Histogram, placing score in the Variable box, to obtain the histogram of Figure 9.5.

The syntax file *POSSKEWDIS9\_5.SPS* appears below:

new file.

input program.

compute prob = -(1/2000).

loop k = 1 to 1000.

compute mean = 8.

compute prob = prob + 1/1000.

compute score = idf.chisq(prob,mean).

leave k, prob, score.

end case.

end loop.

end file.

end input program.

execute.

**(6) THE SPSS SYNTAX PROGRAM, *SAMPDISVER2.SPS,* TO GENERATE THE SET OF 5,000 SAMPLE MEANS USED TO FORM THE SAMPLING DISTRIBUTION OF MEANS GRAPHED AS THE HISTOGRAM OF FIGURE 9.6.**

***SAMPDISVER2.SPS*** reads the 1,000 data points as an external file using the GET FILE command. The 1,000 points originally were generated from the program *POSSKEWDIS9\_5.SPS* and are graphed in Figure 9.5. The external file in which these points reside is located at ‘a:\data\positively skewed population.sav’. Once read, this program, like *SAMPDIS.SPS,* generates a sampling distribution of means. In this case, we have chosen to compute means on 10,000 samples each of size 100 and randomly selected from the positively skewed distribution. This program also incorporates the command for constructing the histogram given Figure 9.6.

Directions for use: Open the file labeled *SAMPDISVER2.SPS* in the SPSS syntax window from the CD that accompanies the text. Click **Run**, **All**. The 5,000 sample means that are generated will appear in the data editor and will also be assigned the file name bootmean.sav. The histogram constructed on these data will appear in the output window and correspond to the histogram of Figure 9.6.

If you wish to work with a different population of values, simply replace ‘a:\data\positively skewed population.sav’ with the location and name of the file in which this new population of values resides. If you change the size of the population, make the appropriate adjustment to the value of nsize. If you wish to change the number of sample means generated, change the value nsamp. If you wish to change the size of each of the samples randomly selected, change the value of samsize.

The syntax file *SAMPDISVER2.SPS* appears below:

DEFINE bootmn (nsamp !tokens(1) / nsize !tokens(1) /samsize !tokens(1)

/ bootvar !tokens(1) / outfile !tokens(1)).

Sort cases by !bootvar.

vector data (!nsize).

compute data ($casenum) = !bootvar.

compute nobreak = 1.

Aggregate outfile \*

/break nobreak

/data1 to !concat(data,!nsize) = max(data1 to !concat(data,!nsize)).

Vector data = data1 to !concat(data,!nsize).

vector tmp(!nsize).

loop #p = 1 to !nsamp.

loop #q = 1 to !nsize.

compute tmp(#q) = 0.

End loop.

Loop #I = 1 to !samsize.

compute id = runk(uniform(!nsize) + 1).

Compute tmp(id) = tmp(id) + 1.

End loop.

Compute mean = 0.

Loop #j = 1 to !nsize.

compute mean = mean + data(#j)\*tmp(#j).

end loop.

Compute mean = mean/!samsize.

\*save to an outfile\*.

Xsave outfile !quote(!outfile)/keep mean.

End loop.

Execute.

!ENDDEFINE.

get file = 'D:\Data CD\positively skewed population.sav' /DROP= prob k mean /KEEP=score.

Bootmn nsamp = 10000 nsize = 1000 samsize = 100 bootvar = x outfile = bootmean.

Get file “bootmean”.

Graph /histogram=mean.

Execute.