

4. Sampling distributions

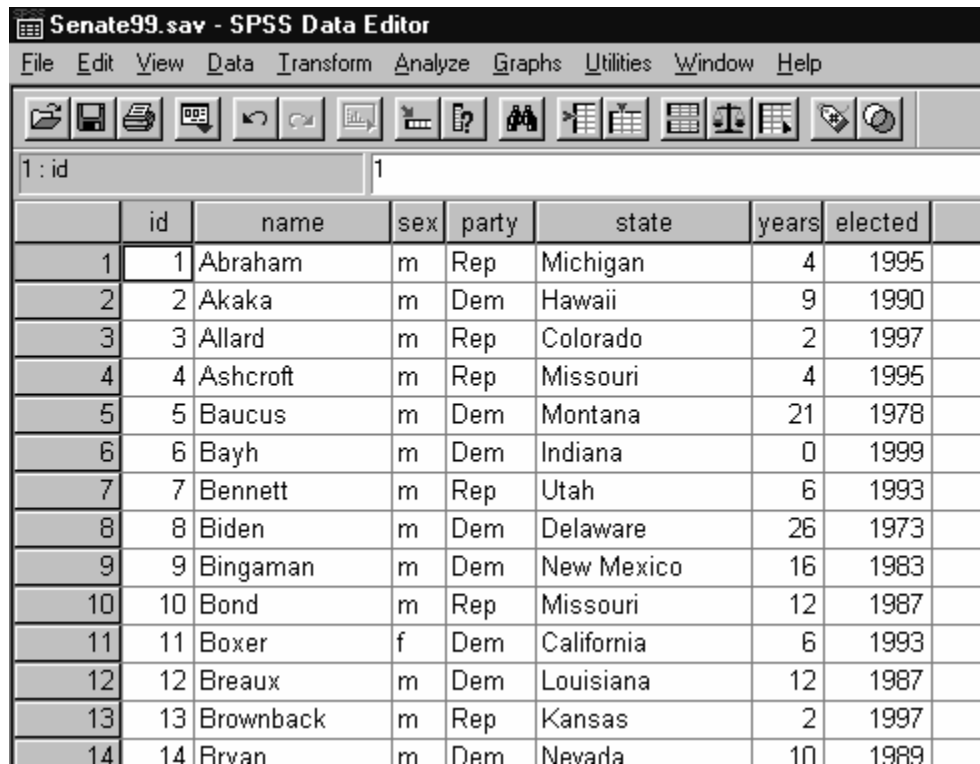
4.1 Selecting a random sample of cases

Topic 4 in the text concerns methods of selecting random samples of cases from some defined population. This section outlines how to select a random sample of cases from a data set in SPSS.

Example 4.1 U.S. Senate. Data on all 100 U.S. Senators in 1999 were collected and entered into SPSS, with the first few cases shown in Figure 4-1.

To select a random sample of 10 cases from the 100 in this data set, follow these steps.

1. Click **Transform**, then click **Random Number Generators**. For SPSS version 16.0, check the **Set Starting Point** box. Make sure “Random” (the default) is selected and then click **OK**. This step tells SPSS to start at a random place in its table of random numbers.
2. Click **Data**, then click **Select Cases**. The “Select Cases” window in Figure 4-2 appears.
3. Click **Random sample of cases**.
4. Click the **Sample** button directly underneath **Random sample of cases**. The “Select Cases: Random Sample” window appears. Figure 4-3 shows how this box will appear after the next step.
5. Click **Exactly**, then enter the size of the sample (10, in this case) to be selected in the first box and the total number of cases in the data set (100, in this case) in the second box. The box should now appear exactly as in Figure 4-3.
6. Click **Continue**.
7. Click **OK**.



	id	name	sex	party	state	years	elected
1	1	Abraham	m	Rep	Michigan	4	1995
2	2	Akaka	m	Dem	Hawaii	9	1990
3	3	Allard	m	Rep	Colorado	2	1997
4	4	Ashcroft	m	Rep	Missouri	4	1995
5	5	Baucus	m	Dem	Montana	21	1978
6	6	Bayh	m	Dem	Indiana	0	1999
7	7	Bennett	m	Rep	Utah	6	1993
8	8	Biden	m	Dem	Delaware	26	1973
9	9	Bingaman	m	Dem	New Mexico	16	1983
10	10	Bond	m	Rep	Missouri	12	1987
11	11	Boxer	f	Dem	California	6	1993
12	12	Breaux	m	Dem	Louisiana	12	1987
13	13	Brownback	m	Rep	Kansas	2	1997
14	14	Bryan	m	Dem	Nevada	10	1989

Figure 4-1

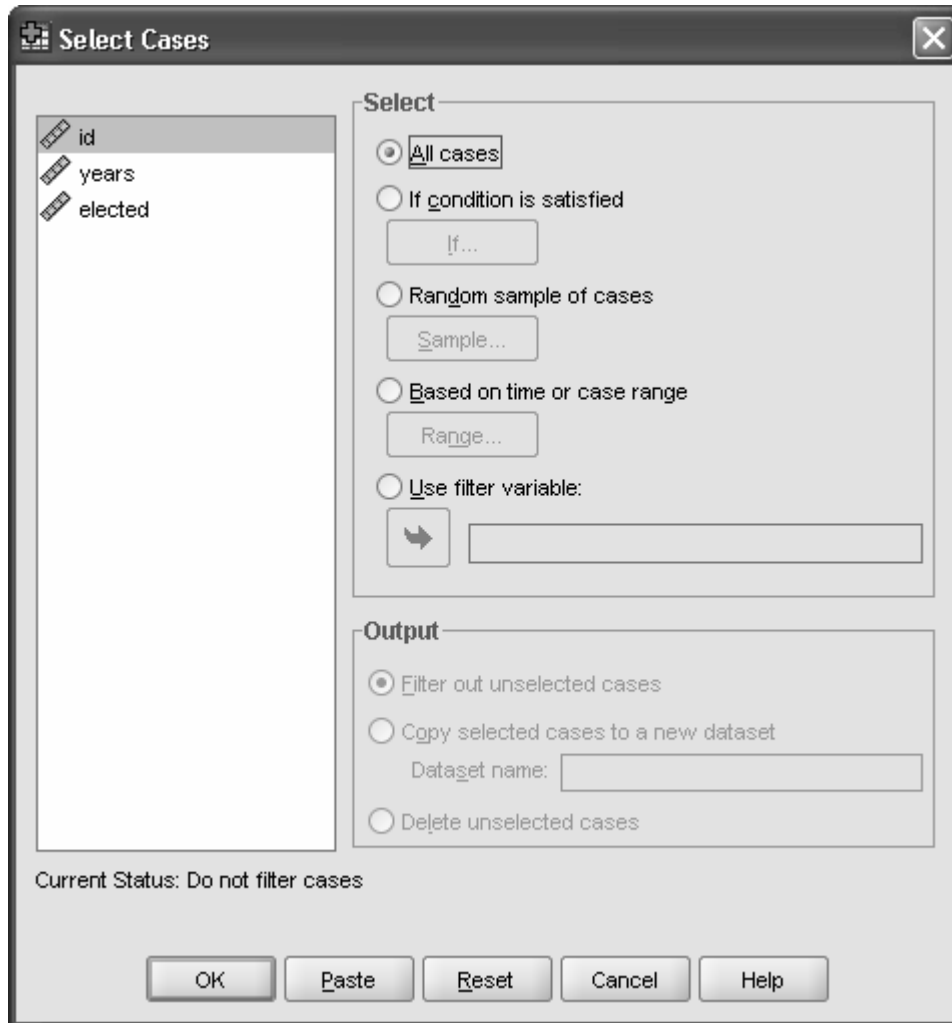


Figure 4-2

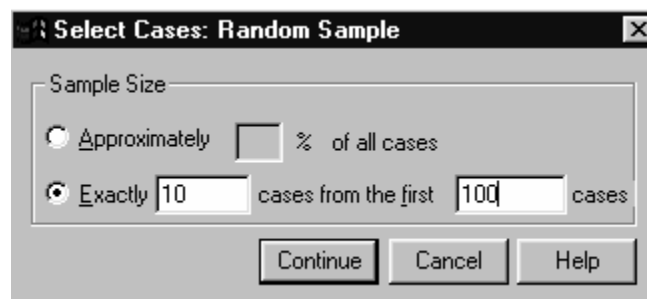


Figure 4-3

The Data Editor will now appear as in Figure 4-4, with a new variable called *filter_\$*. The cases that have been randomly selected have a value of 1 while the rest have a value of 0. Naturally, different cases will be selected every time you run this procedure. Notice also that the cases that have not been selected have a slash through the case number on the left side of the screen.

Senate99.sav - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Window Help

1 : id 1

	id	name	sex	party	state	years	elected	filter_\$
1	1	Abraham	m	Rep	Michigan	4	1995	0
2	2	Akaka	m	Dem	Hawaii	9	1990	0
3	3	Allard	m	Rep	Colorado	2	1997	0
4	4	Ashcroft	m	Rep	Missouri	4	1995	0
5	5	Baucus	m	Dem	Montana	21	1978	0
6	6	Bayh	m	Dem	Indiana	0	1999	0
7	7	Bennett	m	Rep	Utah	6	1993	0
8	8	Biden	m	Dem	Delaware	26	1973	0
9	9	Bingaman	m	Dem	New Mexico	16	1983	1
10	10	Bond	m	Rep	Missouri	12	1987	0
11	11	Boxer	f	Dem	California	6	1993	0
12	12	Breaux	m	Dem	Louisiana	12	1987	1
13	13	Brownback	m	Rep	Kansas	2	1997	0
14	14	Rrvan	m	Dem	Nevada	10	1989	0

Figure 4-4

At this point, any analyses you carry out will automatically be carried out only on the selected cases. For example, to compute the proportion in each party for the 10 selected senators, click **Analyze**, click **Descriptive Statistics**, then click **Frequencies**. Move *party* into the “Variable(s)” box and click **OK**. The result for this example is shown in Table 4-1. Seventy percent of this random sample of 10 senators was Democrat and 30% was Republican.

You can repeat the whole process of taking a random sample and computing summary statistics as many times as you wish. This allows you to examine the sampling distribution of sample proportions, sample means, and other sample statistics.

PARTY

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Dem	7	70.0	70.0	70.0
Rep	3	30.0	30.0	100.0
Total	10	100.0	100.0	

Table 4-1

Undoing the selection of cases

If you want SPSS to go back to using the full data set, click **Data**, then click **Select Cases**. The “Select Cases” window in Figure 4-2 appears. Click **All cases** in the “Select” box, then click **OK**.

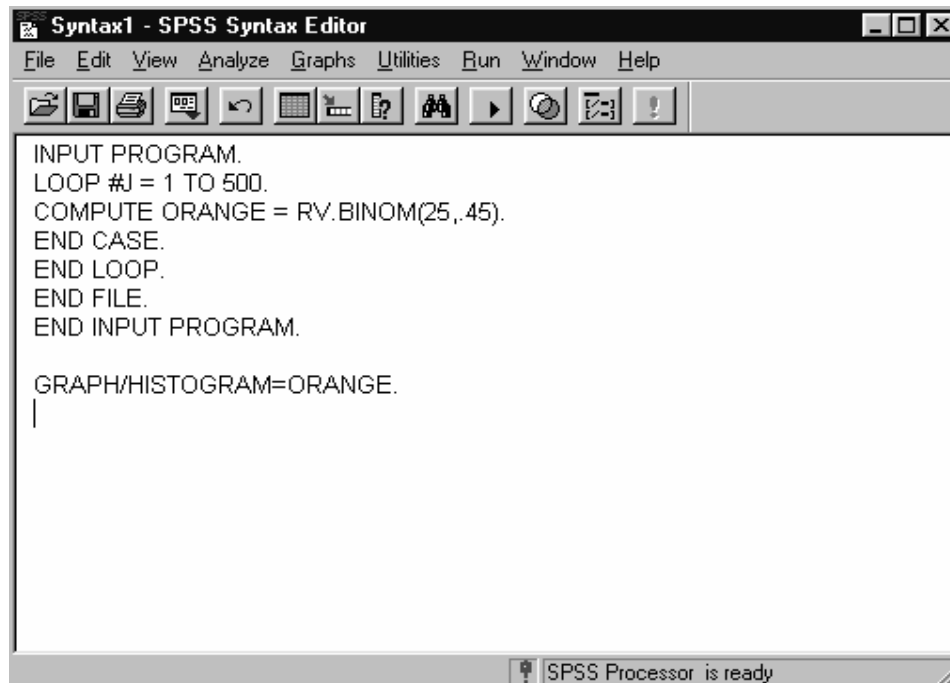
4.2 Simulating binomial data

Some of the activities in Topics 11, 13, and 16 in the text ask you to use the computer to simulate samples from a binary population with population proportion π . To illustrate how to do this, we will use Activity 13-2 for Reese's pieces. **Note that the Student version of SPSS cannot do this simulation because it cannot create and run programs as outlined below.**

Example 4.2 Simulating Reese's Pieces. In Activity 13-2, page 255 of the text, we are told to assume that the proportion of orange Reese's Pieces in the population of all Reese's pieces is $\pi = .45$. We want to simulate 500 random samples of size 25 from this population and calculate the *proportion* of orange pieces in each sample.

The *number* of orange pieces in each sample is what is called a *binomial* random variable. A binomial random variable depends on two quantities: the size of each sample n , and the proportion of orange pieces in the population π . In this example, n is 25, the size of each sample. To have SPSS simulate 500 random samples of size 25 each, you will create a short program in the SPSS Syntax Editor. Follow these steps.

1. Click **Transform**, then click **Random Number Generators**. Make sure "Random" is selected and then click **OK**. This tells SPSS to start at a random place in its table of random numbers.
2. Click **File**, click **New**, and then click **Syntax**. The SPSS Syntax Editor in Figure 4-5 (but without the text) appears.
3. Type the program appearing in Figure 4-5 into the Syntax Editor. Note that each line must end with a period. The first part of the program generates 500 random samples of size $n = 25$ from a binomial distribution with $\pi = .45$ (note the format of the RV.BINOM command). The second part of the program generates a histogram of the resulting values.
4. To run the program, click **Run**, then click **All**. Be sure you have done step 1 before running the program.



```

Syntax1 - SPSS Syntax Editor
File Edit View Analyze Graphs Utilities Run Window Help
[Icons]
INPUT PROGRAM.
LOOP #J = 1 TO 500.
COMPUTE ORANGE = RV.BINOM(25,.45).
END CASE.
END LOOP.
END FILE.
END INPUT PROGRAM.

GRAPH/HISTOGRAM=ORANGE.
|
SPSS Processor is ready
  
```

Figure 4-5

After a few seconds, the histogram will appear in the output window. If you return to the Data Editor, you will see a new column in the data sheet (called *orange*) with numbers like 10.00, 11.00, 8.00, etc. in it. These represent the number of orange pieces in each of the 500 samples you generated.

You can summarize the variable *orange* using methods in the previous chapters of this manual. For example, to create a frequency distribution (displaying the number of times there were 0, 1, 2, 3, etc. orange pieces in a sample) and to calculate the mean and standard deviation of the counts, follow these steps (see Section 2.2 Descriptive statistics for quantitative variables): click **Analyze**, then click **Descriptive Statistics**, then click **Frequencies**. Enter *orange* as the variable and make sure **Display frequency tables** is checked. Click on the **Statistics** button and make sure **Mean** and **Std. Deviation** are checked. Then click **Continue**, and click **OK**. The resulting tables are shown as Table 4-2 and Table 4-3. Of course, if you were to run this simulation, you would not get exactly the same results.

Statistics

ORANGE		
N	Valid	500
	Missing	0
Mean		11.2740
Std. Deviation		2.4213

Table 4-2

ORANGE

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 4.00	1	.2	.2	.2
5.00	6	1.2	1.2	1.4
6.00	4	.8	.8	2.2
7.00	19	3.8	3.8	6.0
8.00	40	8.0	8.0	14.0
9.00	48	9.6	9.6	23.6
10.00	59	11.8	11.8	35.4
11.00	83	16.6	16.6	52.0
12.00	83	16.6	16.6	68.6
13.00	67	13.4	13.4	82.0
14.00	48	9.6	9.6	91.6
15.00	28	5.6	5.6	97.2
16.00	11	2.2	2.2	99.4
18.00	1	.2	.2	99.6
19.00	2	.4	.4	100.0
Total	500	100.0	100.0	

Table 4-3

To calculate the *proportion* of orange pieces in each of the 500 simulated samples, go back to the Data Editor and follow these steps (see Section 1.14, page 12): click **Transform**, then click **Compute Variable**. Enter a variable name for the Target Variable (e.g., *prop*) and type the expression *orange/25*

in the “Numeric Expression” box (you divide by whatever your n is, 25 in this case). Now you can create a frequency table for this variable, if desired, compute summary statistics, a histogram, etc.

The program in the Syntax Editor can be saved by clicking **File** and **Save** in the Syntax Editor window. The program will be saved as a .SPS file that can be retrieved, edited and rerun later. To open a .SPS file, from the Data Editor, click **File**, click **Open**, and click **Syntax**. Be sure to again set the random number seed (step 1) when running the program later.