sampling be an appropriate technique (assuming that no list was available)? Describe in detail how you would construct a cluster sample.

6.2 This exercise is extremely tedious and hardly ever works out the way it ought to, mostly because not many people have the patience to draw an "infinite" number of even very small samples. However, if you want a more concrete and tangible understanding of sampling distributions and the two theorems presented in this chapter, then this exercise may have a significant payoff. At the end of this problem are listed the ages of a population of college students ($N = 50$). By a random method (such as a table of random numbers), draw at least 50 samples of size 2 (that is, 50 pairs of cases), compute a mean for each sample, and plot the means on a frequency polygon. (Incidentally, this exercise will work better if you draw 100 or 200 samples and/or use larger samples than $N = 2$.)

a. The curve you've just produced is a sampling distribution. Observe its shape; after 50 samples, it should be approaching normality. What is your estimate of the population mean (μ) based on the shape of the curve?

b. Calculate the mean of the sampling distribution ($\mu_{\bar{X}}$). Be careful to do this by summing the sample means (not the scores) and dividing by the number of samples you've drawn. Now compute the population mean (μ). These two means should be very close in value because $\mu_{\bar{X}} = \mu$ by the Central Limit Theorem.

c. Calculate the standard deviation of the sampling distribution (use the means as scores) and the standard deviation of the population. Compare these two values. You should find that $\sigma_{\bar{X}} = \sigma/\sqrt{N}$.

d. If none of the preceding exercises turned out as they should have, it is for one or more of the following reasons:

1. You didn't take enough samples. You may need as many as 100 or 200 (or more) samples to see the curve begin to look "normal."
2. Sample size (2) is too small. An N of 5 or 10 would work much better.
3. Your sampling method is not truly random and/or the population is not arranged in random fashion.

17	20	20	19	20
18	21	19	20	19
19	22	19	23	19
20	23	18	20	20
22	19	19	20	20
23	17	18	21	20
20	18	20	19	20
22	17	21	21	21
21	20	20	20	22
18	21	20	22	21

Using SPSS for Windows to Draw Random Samples

DEMONSTRATION 6.1 Estimating Average Age

SPSS for Windows includes a procedure for drawing random samples from a database. We can use this procedure to illustrate some points about sampling and to convince the skeptics in the crowd that properly selected samples will produce statistics that are close approximations of the corresponding population values or parameters. For purposes of this demonstration, the 2006 GSS sample will be treated as a population and its characteristics will be treated as parameters.

The following instructions will calculate a mean for *age* for three random samples of different sizes drawn from the 2006 GSS sample. The actual average age of the sample (which will be the parameter or μ) is 46.88 (see Demonstration 5.1). The samples are roughly 10%, 25%, and 50% of the population size, and the program selects them by an EPSEM procedure. Therefore, these samples may be considered "simple random samples."

As a part of this procedure we also request the "standard error of the mean," or S.E. MEAN. This is the standard deviation of the sampling distribution ($\sigma_{\bar{X}}$) for a sample of this size. This statistic will be of interest because we can expect our sample means to be within this distance of the population value or parameter.

With the 2006 GSS loaded, click **Data** from the menu bar of the Data Editor and then click **Select Cases.** The **Select Cases** window appears and presents a number of different options. To select random samples, click the button next to "Random sample of cases" and then click on the **Sample** button. The **Select Cases: Random Sample** window will open. We can specify the size of the sample in two different ways. If we use the first option, we can specify that the sample will include a certain percentage of cases in the database. The second option allows us to specify the exact number of cases in the sample. Let's use the first option and request a 10% sample by typing 10 into the box on the first line. Click **Continue,** and then click **OK** on the **Select Cases** window. The sample will be selected and can now be processed.

To find the mean age for the 10% sample, click **Analyze, Descriptive Statistics**, and then **Descriptives.** Find *age* in the variable list and transfer it to the **Variable(s):** window. On the **Descriptives** menu, click the **Options** button and select S.E. MEAN in addition to the usual statistics. Click **Continue** and then **OK,** and the requested statistics will appear in the output window.

Now, to produce a 25% sample, return to the **Select Cases** window by clicking **Data** and **Select Cases.** Click the **Reset** button at the bottom of the window and then click **OK** and the full data set (N = 1417) will be restored. Repeat the procedure we followed for selecting the 10% sample. Click the button next to "Random sample of cases" and then click on the **Sample** button. The **Select Cases: Random Sample** window will open. Request a 25% sample by typing 25 in the box, click **Continue** and **OK,** and the new sample will be selected.

Run the **Descriptives** procedure for the 25% sample (don't forget S.E. MEAN) and note the results. Finally, repeat these steps for a 50% sample. The results are summarized here:

1 Sample %	2 Sample Size	3 Sample Mean	4 Standard Error	5 Sample Mean ± Standard Error	6 Sample Mean − Population Mean
10%	133	47.61	1.46	46.15–49.07	0.73
25%	361	46.89	0.91	45.98–47.80	0.01
50%	653	46.99	0.67	46.32–47.66	0.11

Let's look at this table column by column. The first column lists the sample percent and column 2 lists the actual number of cases in each sample. Sample means and standard errors (or standard deviations of the sampling distribution) are listed in columns 3 and 4. Look at column 4 and note that standard error decreases as sample size increases. This should reinforce the commonsense notion that larger samples will provide more accurate estimates of population values.

To produce column 5, I subtracted and then added the value of the standard error to the mean of each sample. Note that all three intervals listed in column 5 include the value of the population mean (46.88). The important point here is that all three sample means were close to (within 1 standard error of) the population mean. Finally, column 6 shows the distance between the sample means and the population mean. The mean of the smallest (10%) sample is furthest from the population mean. The other two sample means are quite close to the population mean, and the mean of the 25% sample is almost exactly equal to the population value. We would expect the mean of the largest (50%) sample to be closest to the population mean, but remember that we are dealing with chance and probabilities here. At any rate, the most important point is that all three sample means are within 1 standard error of the population mean.

This demonstration should reinforce one of the main points of this chapter: Statistics calculated on samples that have been selected according to the principle of EPSEM will (almost always) be reasonable approximations of their population counterparts.

Exercise

6.1 Following the procedures in Demonstration 6.1, select three samples from the 2006 GSS database: 15%, 30%, and 60%. Get descriptive statistics for *age* (don't forget to get the standard error), and use the results to complete the following table:

1 Sample %	2 Sample Size	3 Sample Mean	4 Standard Error	5 Sample Mean ± Standard Errorfdas	6 Sample Mean Mean-Population
15%	____	____	____	____	____
30%	____	____	____	____	____
60%	____	____	____	____	____

Summarize these results. What happens to standard error as sample size increases? Why? How accurate are the estimates (sample means)? Are all sample means within a standard error of 46.88? How does the accuracy of the estimates change as sample size changes?