CHAPTER 6—ANSWERS TO EXERCISES

1.

   a. Although there are problems with the collection of data from all Americans, the census is assumed to be complete, so the mean age would be a parameter.

   b. A statistic because it is estimated from a sample.

   c. A statistic because it is estimated from a sample.

   d. A parameter because the school has information on all employees.

   e. A parameter because the school would have information on all its students.

2.

   a. First, it is not clear what the population is. Are the population subscribers to the newspaper? Readers of the newspaper? Or some other set of people? Even given this uncertainty, the letters to the editor would only be a random sample of a population if, clearly, they came randomly from that population. But there is no reason to assume this to be the case. People with stronger opinions, who can write reasonably well, and who have the time to write are more likely to write to the editor. Since these characteristics are not distributed randomly throughout the adult population, it is improbable that the letters are a random sample.

   b. The mayor might consider forming a coalition to randomly sample landline and cell phone numbers of city residents.

3.

   a. Assuming that the population is defined as all persons shopping at that shopping mall that day of the week, she is selecting a systematic random sample. A more precise definition might limit it to all persons passing by the department store at the mall that day.

   b. This is a stratified sample because voters were first grouped by county, and unless the counties have the same number of voters, it is a disproportionate stratified sample because the same number is chosen from each county. We can assume that it was a probability sample, but we are not told exactly how the 50 voters were chosen from the lists. However, assuming that the population is defined as all Americans, this sort of sampling technique would qualify as nonprobability sampling.

   c. This is neither a simple random sample nor a systematic random sample. It might be thought of as a sample stratified on last name, but even then, choosing the first 20 names is not a random selection process.

   d. This is not a probability sample. Instead, it is a purposive sample chosen to represent a cross-section of the population in New York City.

4.

a. There are 120 students, of whom 57 are juniors. Juniors represent 57/120 = 47.5% of the class. The probability of choosing a junior is .475.

b. The probability that the student will be a freshman is 7/120 = .058.

c. The proportion of seniors in the class is 34/120 = .283. The proportion of sophomores is 22/120 = .183. Then the number of students to be chosen at each class level is as follows:

Freshman	.058(30) = 1.74 or about 2
Sophomore	.183(30) = 5.49 or about 5
Junior	.475(30) = 14.25 or about 14
Senior	.283(30) = 8.49 or about 8

Due to rounding, the total comes out to 29 students, not 30. This is typical of real-world applications of sampling. If we wanted exactly 30 students, we could choose one more sophomore or senior.

d. For a disproportionate sample, we choose the same number of students from each class level, so we would have five freshmen (and five sophomores, five juniors, and five seniors).

5. The relationship between the standard error and the standard deviation is $\sigma_{\bar{Y}} = \sigma / \sqrt{N}$. Since σ is divided by, $\sqrt{N}$, $\sigma_{\bar{Y}} = \sigma / \sqrt{N}$ must always be smaller than σ, except in the trivial case where $N = 1$. Theoretically, the dispersion of the mean must be less than the dispersion of the raw scores. This implies that the standard error of the mean is less than the standard deviation.

6.

a. The standard error of the mean is proportional to $1 / \sqrt{N}$. The standard error of the mean is

$$\sigma_{\bar{Y}} = \frac{\sigma}{\sqrt{100}} = \frac{\sigma}{10} \text{ for a sample size of 100.}$$

$$\sigma_{\bar{Y}} = \frac{\sigma}{\sqrt{1{,}600}} = \frac{\sigma}{40} \text{ for a sample size of 1,600.}$$

Clearly, $\sigma_{\bar{Y}} = \sigma / \sqrt{N}$ is smaller by a factor of 4 when sample size increases to 1,600.

b.

$$\sigma_{\bar{Y}} = \frac{\sigma}{\sqrt{300}} = \frac{\sigma}{17.32}$$

$$\sigma_{\bar{Y}} = \frac{\sigma}{\sqrt{150}} = \frac{\sigma}{12.25}$$

So when sample size decreases, standard error of the mean increases by

$$\frac{17.32}{12.25} = 1.41$$

c. Assume an initial sample size of 100. Any initial value will suffice. Then,

$$\sigma_{\bar{Y}} = \frac{\sigma}{\sqrt{100}} = \frac{\sigma}{10}$$

$$\sigma_{\bar{Y}} = \frac{\sigma}{\sqrt{4(100)}} = \frac{\sigma}{20}$$

So standard error decreases by a factor of 2, which is the square root of 4.

7.

a. These polls are definitely not probability samples. No sampling is done by the television station to choose who calls the 800 number.

b. The population is all those people who watch the television channel and see the 800 number advertised.

8.

a. The mean number of active military personnel per region in 2009 was

$$\bar{Y} = \frac{1{,}082{,}228}{9} = 120{,}247.6$$

with a standard deviation of 119,819.

b. 10 sample means calculated from samples of size 3:

223,410.7

52,547.0

84,785.0

165,595.3

53,129.3

142,018.0

88,199.0

193,251.3

41,593.0

89,428.0

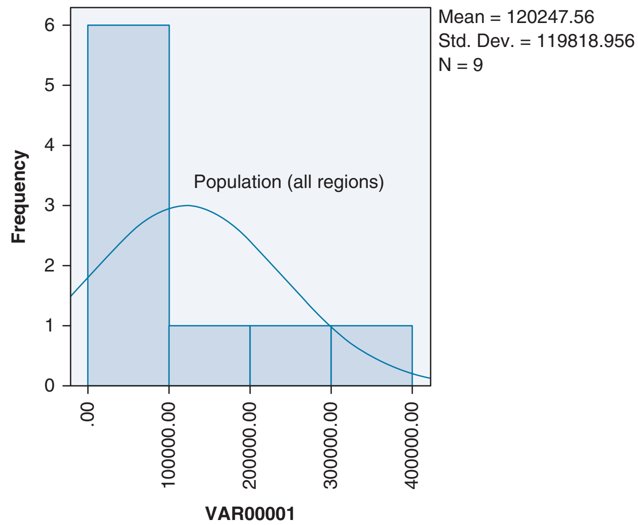
c. The mean of these 10 means is 113,395.66. Right away we notice that the population mean and the mean of the sampling distribution are somewhat close, a feature that we should come to expect given the fact that $\mu = \mu_{\bar{Y}}$.

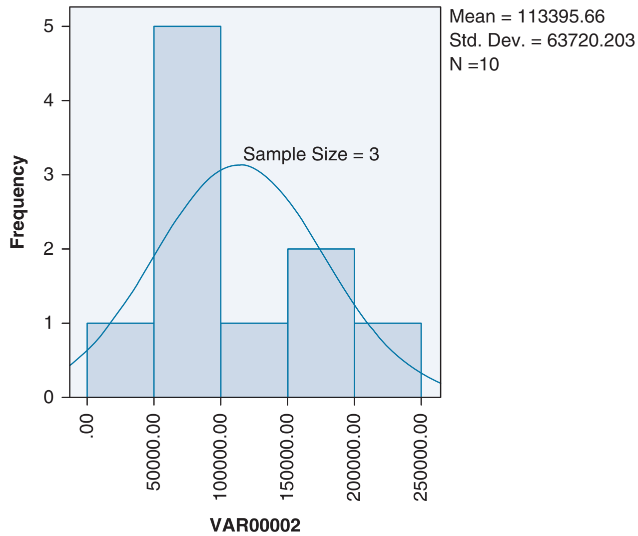
d. The standard deviation for all regions is 119,819. The standard error is calculated using the following equation:

$$\sigma_{\bar{Y}} = \frac{\sigma}{\sqrt{N}} = \frac{119{,}819}{\sqrt{3}} = 69{,}177.53$$

We know that the value of the standard error will always be less than the value of the standard deviation in the population.

e. The population distribution is positively skewed and not close to normal. Since a very small sample size is used in this problem, the histogram for the samples of size 3 does not look normal. The distributions appear unimodal. The fact that the sample distribution of the means tends toward normality because of the central limit theorem would become even more apparent if we took samples of size 5 or 6.





f. We treated the distribution for all regions as the population distribution.

9.

a. This is not a random sample. The students eating lunch on Tuesday are not necessarily representative of all students at the school, and you have no way of calculating the probability of inclusion of any student. Many students might, for example, rarely eat lunch at the cafeteria and, therefore, have no chance of being represented in your sample. The fact that you selected *all* the students eating lunch on Tuesday makes your selection appear to be a census of a population, but that isn't true either unless all the students ate at the cafeteria on Tuesday.

b. This is a systematic random sample because names are drawn systematically from the list of all enrolled students.

c. This would seem to be a systematic random sample as in (b), but it suffers from the same type of defect as the cafeteria sample. Unless all students pass by the students union, using that location as a selection criterion means that some students have no chance of being selected (but you don't know which ones). Samples are often drawn this way in shopping malls by choosing a central location from which to draw the sample. It is reasonable to assume that a sufficiently representative mix of shoppers will pass by a central location during any one period.

d. The second procedure (selecting every 10th student from the list of all enrolled students) is the best option because it uses a random sampling method.

10.

The first five respondents chosen are respondents numbered 15,179, 18,602, 18,663, 18,594, and 18,912. All other values in Appendix A are higher than 21,473; therefore, they do not represent valid cases for inclusion in the sample.

11.

a. Mean = 5.3 (53/10); standard deviation = 3.27.

b. Here are 10 means from random samples of size 3: 6.33, 5.67, 3.33, 5.00, 7.33, 2.33, 6.00, 6.33, 7.00, 3.00.

c. The mean of these 10 sample means is 5.23. The standard deviation is 1.76. The mean of the sample means is very close to the mean for the population. The standard deviation of the sample means is much less than the standard deviation for the population. The standard deviation of the means from the samples is an estimate of the standard error of the mean we would find from one random sample of size 3.

12.

a. The standard error is calculated as follows:

$$\sigma_{\overline{Y}} = \frac{\sigma}{\sqrt{N}} = \frac{226.83}{\sqrt{12}} = 65.48$$

This value represents the average standard deviation of any sample mean from the mean of means. Accordingly, it may also be referred to as the standard deviation of the sampling distribution.

b. With the exception of cases where $N = 1$, the standard error will always be less in value than the standard deviation of the population. This is expressed by the formula

$$\sigma_{\overline{Y}} = \frac{\sigma}{\sqrt{N}}$$

The shape of the sampling distribution is normal; thus, even when working with a skewed distribution, we know that the sampling distribution is normal. Suggestions for reducing the sampling error include increasing the sample size.

SPSS SOLUTIONS

The purpose of this exercise is to demonstrate the central limit theorem, that is, as sample size increases, the standard error decreases.

25% Sample	N	Mean	Standard Error
MAEDUC	362	11.48	4.026
PAEDUC	291	11.78	4.102

50% Sample	N	Mean	Standard Error
MAEDUC	658	11.58	3.767
PAEDUC	552	11.59	4.279

75% Sample	N	Mean	Standard Error
MAEDUC	1023	11.60	3.915
PAEDUC	823	11.69	4.276

100% Sample	N	Mean	Standard Error
MAEDUC	1363	11.55	3.924
PAEDUC	1099	11.61	4.270