7.9 CONFIDENCE INTERVALS FOR PARAMETERS OF FINITE POPULATIONS (OPTIONAL)

It is best to use the confidence intervals presented in Sections 7.1 through 7.4 when the sampled population is either infinite or finite and *much larger than* (say, at least 20 times as large as) the sample. Although these previously discussed intervals are sometimes used when a finite population is not much larger than the sample, better methods exist for handling such situations. We present these methods in this section.

As we have explained, we often wish to estimate a population mean. Sometimes we also wish to estimate a **population total**.

A **population total** is the sum of the values of all of the population measurements.

For example, companies in financial trouble have sometimes falsified their accounts receivable invoices in order to mislead stockholders. For this reason, independent auditors are often asked to estimate a company's true total sales for a given period. The auditor randomly selects a sample of invoices from the population of all invoices and then independently determines the actual amount of each sale by contacting the purchasers. The sample results are used to estimate the company's total sales, and this estimate can then be compared with the total sales reported by the company.

In order to estimate a population total, which we denote as τ (tau), we note that the population mean μ is the population total divided by the number, N, of population measurements. That is, we have $\mu = \tau/N$, which implies that $\tau = N\mu$. Then, because a point estimate of the population mean μ is the sample mean $\bar{x}$, we have the following:

A **point estimate of a population total τ** is $N\bar{x}$, where N is the size of the population.

Example 7.17 Estimating True Total Sales

A company sells and installs satellite dishes and receivers for both private individuals and commercial establishments (bars, restaurants, and so forth). The company accumulated 2,418 sales invoices last year. The total of the sales amounts listed on these invoices (that is, the total sales claimed by the company) is \$5,127,492.17. In order to estimate the true total sales, τ, for last year, an independent auditor randomly selects 242 of the invoices and determines the actual sales amounts by contacting the purchasers. When the sales amounts are averaged, the mean of the actual sales amounts for the 242 sampled invoices is $\bar{x} = \$1,843.93$. This says that a point estimate of the true total sales τ is

$$N\bar{x} = 2,418(\$1,843.93) = \$4,458,622.74.$$

This point estimate is considerably lower than the claimed total sales of \$5,127,492.17. However, we cannot expect the point estimate of τ to exactly equal the true total sales, so we need to calculate a confidence interval for τ before drawing any unwarranted conclusions.

In order to find a confidence interval for the mean and total of a finite population, we consider the sampling distribution of the sample mean $\bar{x}$. It can be shown that if we randomly select a large sample of n measurements without replacement from a finite population of N measurements, the sampling distribution of $\bar{x}$ is approximately normal with mean $\mu_{\bar{x}} = \mu$ and standard deviation

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}.$$

It can also be shown that the appropriate point estimate of $\sigma_{\bar{x}}$ is $(s/\sqrt{n})(\sqrt{(N-n)/N})$, where s is the sample standard deviation. This point estimate of $\sigma_{\bar{x}}$ is used in the confidence intervals for μ and τ, which we summarize as follows:

Confidence Intervals for the Population Mean and Population Total for a Finite Population

Suppose we randomly select a sample of n measurements *without replacement from a finite population of N measurements*. Then, if n is large (say, at least 30),

1 A **100(1 − α) percent confidence interval for the population mean μ is**

$$\left[\bar{x} \pm z_{\alpha/2}\frac{s}{\sqrt{n}}\sqrt{\frac{N-n}{N}}\right].$$

2 A **100(1 − α) percent confidence interval for the population total τ** is found by multiplying the lower and upper limits of the 100(1 − α) percent confidence interval for μ by N.

The quantity $\sqrt{(N-n)/N}$ in the confidence intervals for μ and τ is called the **finite population correction**. If the population size N is much larger than (say, at least 20 times as large as) the sample size n, then the finite population correction is approximately equal to 1. For example, if we randomly select (without replacement) a sample of 1,000 from a population of 1 million, then the finite population correction is $\sqrt{(1{,}000{,}000 - 1{,}000)/1{,}000{,}000} = 0.9995$. In such a case, many people believe it is not necessary to include the finite population correction in the confidence interval calculations. This is because the correction is not far enough below 1 to meaningfully shorten the confidence intervals for μ and τ. However, *if the population size N is not much larger than the sample size n (say, if n is more than 5 percent of N), then the finite population correction is substantially less than 1 and should be included* in the confidence interval calculations.

Example 7.18 Overstating Sales Invoices?

Recall that the satellite dish dealer claims that its total sales, τ, for last year were \$5,127,492.17. Since the company accumulated 2,418 invoices during last year, the company is claiming that μ, the mean sales amount per invoice, is \$5,127,492.17/2,418 = \$2,120.55. Suppose, when the independent auditor randomly selects a sample of $n = 242$ invoices, that the mean and standard deviation of the actual sales amounts for these invoices are $\bar{x} = 1{,}843.93$ and $s = 516.42$. Here the sample size $n = 242$ is $(242/2{,}418)100 = 10.008$ percent of the population size $N = 2{,}418$. Because n is more than 5 percent of N, we should include the finite population correction in our confidence interval calculations. It follows that a 95 percent confidence interval for the mean sales amount μ per invoice is

$$\left[\bar{x} \pm z_{0.025}\frac{s}{\sqrt{n}}\sqrt{\frac{N-n}{N}}\right] = \left[1{,}843.93 \pm 1.96\left(\frac{516.42}{\sqrt{242}}\right)\sqrt{\frac{2{,}418-242}{2{,}418}}\right]$$
$$= [1{,}843.93 \pm 61.723812]$$
$$= [1{,}782.21,\ 1{,}905.65].$$

The upper limit of this interval is less than the mean amount of \$2,120.55 claimed by the company, so we have strong evidence that the company is overstating its mean sales per invoice for last year. A 95 percent confidence interval for the total sales, τ, last year is found by multiplying the lower and upper limits of the 95 percent confidence interval for μ by $N = 2{,}418$.

Therefore, this interval is [1,782.21(2,418), 1,905.65(2,418)], or [4,309,383.78, 4,607,861.70]. Because the upper limit of this interval is more than \$500,000 below the total sales amount of \$5,127,492.17 claimed by the company, we have strong evidence that the satellite dealer is substantially overstating its total sales for last year.

We sometimes estimate the total number, τ, of population units that fall into a particular category. For instance, the auditor of Examples 7.17 and 7.18 might wish to estimate the total number of the 2,418 invoices with incorrect sales amounts. Here the proportion, p, of the population units that fall into a particular category is the total number, τ, of population units that fall into the category divided by the number, N, of population units. That is, $p = \tau/N$, which implies that $\tau = Np$. Therefore, since a point estimate of the population proportion p is the sample proportion $\hat{p}$, a point estimate of the population total τ is $N\hat{p}$. For example, suppose that 34 of the 242 sampled invoices have incorrect sales amounts. Because the sample proportion is $\hat{p} = 34/242 = 0.1405$, a point estimate of the total number of the 2,418 invoices that have incorrect sales amounts is

$$N\hat{p} = 2{,}418(0.1405) = 339.729.$$

We now summarize how to find confidence intervals for p and τ.

Confidence Intervals for the Proportion of and Total Number of Units in a Category When Sampling a Finite Population

Suppose that we randomly select a sample of n units *without replacement from a finite population of N units*. Then, if n is large, we have the following:

1 A **100(1 − α) percent confidence interval for the population proportion p** is

$$\left[\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n-1}\left(\frac{N-n}{N}\right)}\right].$$

2 A **100(1 − α) percent confidence interval for the population total τ** is found by multiplying the lower and upper limits of the 100(1 − α) percent confidence interval for p by N.

Example 7.19 Proportion of Incorrect Invoices

Recall that in Examples 7.17 and 7.18 we found that 34 of the 242 sampled invoices have incorrect sales amounts. Since $\hat{p} = 34/242 = 0.1405$, a 95 percent confidence interval for the proportion of the 2,418 invoices that have incorrect sales amounts is

$$\left[\hat{p} \pm z_{0.025}\sqrt{\frac{\hat{p}(1-\hat{p})}{n-1}\left(\frac{N-n}{N}\right)}\right] = \left[0.1405 \pm 1.96\sqrt{\frac{(0.1405)(0.8595)}{241}\left(\frac{2{,}418-242}{2{,}418}\right)}\right]$$
$$= [0.1405 \pm 0.0416208]$$
$$= [0.0989, 0.1821].$$

This interval says we are 95 percent confident that between 9.89 percent and 18.21 percent of the invoices have incorrect sales amounts. A 95 percent confidence interval for the total number of the 2,418 invoices that have incorrect sales amounts is found by multiplying the lower and upper limits of the 95 percent confidence interval for p by $N = 2{,}418$. Therefore, this interval is [0.0989(2,418), 0.1821(2,418)], or [239.14, 440.32], and we are 95 percent confident that between (roughly) 239 and 440 of the 2,418 invoices have incorrect sales amounts.

Finally, we can determine the sample size that is needed to make the margin of error in a confidence interval for μ, p, or τ equal to a desired size E by setting the appropriate margin of error formula equal to E and by solving the resulting equation for the sample size n (the procedure is the same as illustrated in Sections 7.3 and 7.4). Exercise 7.119 gives the reader an opportunity to use the sample size formulas that are obtained.

Exercises for Section 7.9

CONCEPTS

7.113 Define a population total. Give an example of a population total that will interest you in your career when you graduate from university.

7.114 Explain why the finite population correction $\sqrt{(N-n)/N}$ is unnecessary when the population is at least 20 times as large as the sample. Give an example using numbers.

METHODS AND APPLICATIONS

7.115 A retailer that sells home entertainment systems accumulated 10,451 sales invoices during the previous year. The total of the sales amounts listed on these invoices (that is, the total sales claimed by the company) is \$6,384,675. In order to estimate the true total sales for last year, an independent auditor randomly selects 350 of the invoices and determines the actual sales amounts by contacting the purchasers. The mean and the standard deviation of the 350 sampled sales amounts are $\bar{x} = \$532$ and $s = \$168$.

a. Find a 95 percent confidence interval for μ, the true mean sales amount per invoice on the 10,451 invoices.

b. Find a point estimate of and a 95 percent confidence interval for τ, the true total sales for the previous year.

c. What does this interval say about the company's claim that the true total sales were \$6,384,675? Explain.

7.116 A company's manager is considering simplifying a travel voucher form. In order to assess the costs associated with erroneous travel vouchers, the manager must estimate the total number of such vouchers that were filled out incorrectly in the last month. In a random sample of 100 vouchers drawn without replacement from the 1,323 travel vouchers submitted in the last month, 31 vouchers were filled out incorrectly.

a. Find a point estimate of and a 95 percent confidence interval for the true proportion of travel vouchers that were filled out incorrectly in the last month.

b. Find a point estimate of and a 95 percent confidence interval for the total number of travel vouchers that were filled out incorrectly in the last month.

c. If it costs the company \$10 to correct an erroneous travel voucher, find a reasonable estimate of the minimum cost of correcting all of last month's erroneous travel vouchers. Would it be worthwhile to spend \$5,000 to design a simplified travel voucher that could be used for at least a year?

7.117 A personnel manager is estimating the total number of person-days lost to unexcused absences by hourly workers in the last year. In a random sample of 50 employees drawn without replacement from the 687 hourly workers at the company, records show that the 50 sampled workers had an average of $\bar{x} = 4.3$ days of unexcused absences over the past year with a standard deviation of $s = 1.26$.

a. Find a point estimate of and a 95 percent confidence interval for the total number of unexcused absences by hourly workers in the last year.

b. Can the personnel manager be 95 percent confident that more than 2,500 person-days were lost to unexcused absences last year? Can the manager be 95 percent confident that more than 3,000 person-days were lost to unexcused absences last year? Explain.

7.118 An auditor randomly samples 32 accounts receivable without replacement from a firm's 600 accounts and checks to verify that all documents for the accounts comply with company procedures. Ten of the 32 accounts are found to have documents not in compliance. Find a point estimate of and a 95 percent confidence interval for the total number of accounts with documents that do not comply with company procedures.

7.119 SAMPLE SIZES WHEN SAMPLING FINITE POPULATIONS

a. *Estimating* μ and τ
Consider randomly selecting a sample of n measurements without replacement from a finite population consisting of N measurements and having variance σ^2. Also consider the sample size given by the formula

$$n = \frac{N\sigma^2}{(N-1)D + \sigma^2}.$$

Then it can be shown that this sample size makes the margin of error in a $100(1-\alpha)$ percent confidence interval for μ equal to E if D is set equal to $(E/z_{\alpha/2})^2$. It can also be shown that this sample size makes the margin of error in a $100(1-\alpha)$ percent confidence interval for τ equal to E if D is set equal to $[E/(z_{\alpha/2}N)]^2$. Now consider Exercise 7.117. Using $s^2 = (1.26)^2$, or 1.5876, as an estimate of σ^2, determine the sample size that makes the margin of error in a 95 percent confidence interval for the *total number* of person-days lost to unexcused absences last year equal to 100 days.

b. *Estimating p and τ*
Consider randomly selecting a sample of n units without replacement from a finite population consisting of N units and having a proportion p of these units fall into a particular category. Also consider the sample size given by the formula

$$n = \frac{Np(1-p)}{(N-1)D + p(1-p)}.$$

It can be shown that this sample size makes the margin of error in a $100(1-\alpha)$ percent confidence interval for p equal to E if D is set equal to $(E/z_{\alpha/2})^2$. It can also be shown that this sample size makes the margin of error in a $100(1-\alpha)$ percent confidence interval for τ equal to E if D is set equal to $[E/(z_{\alpha/2}N)]^2$. Now consider Exercise 7.116. Using $\hat{p} = 0.31$ as an estimate of p, determine the sample size that makes the margin of error in a 95 percent confidence interval for the *proportion* of the 1,323 vouchers that were filled out incorrectly equal to 0.04.