7.3.1 The Expectation and Variance of the Sample Mean

We will denote the sample size by n (n is less than N) and the values of the sample members by $X_1, X_2, \ldots, X_n$. It is important to realize that each X_i is a random variable. In particular, X_i is not the same as x_i: X_i is the value of the ith member of the sample, which is random and x_i is that of the ith member of the population, which is fixed.

We will consider the **sample mean,**

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

as an estimate of the population mean. As an estimate of the population total, we will consider

$$T = N\overline{X}$$

Properties of T will follow readily from those of $\overline{X}$. Since each X_i is a random variable, so is the sample mean; its probability distribution is called its **sampling distribution.** In general, any numerical value, or statistic, computed from a random sample is a random variable and has an associated sampling distribution. The sampling distribution of $\overline{X}$ determines how accurately $\overline{X}$ estimates μ; roughly speaking, the more tightly the sampling distribution is centered on μ, the better the estimate.

EXAMPLE **A** To illustrate the concept of a sampling distribution, let us look again at the population of 393 hospitals. In practice, of course, the population would not be known, and only one sample would be drawn. For pedagogical purposes here, we can consider the sampling distribution of the sample mean from this known population. Suppose, for example, that we want to find the sampling distribution of the mean of a sample of size 16. In principle, we could form all $\binom{393}{16}$ samples and compute the mean of each one—this would gi[illegible] sampling distribution. But because the number of such samples is of the order 10[illegible], this is clearly not practical. We will thus employ a technique known as **simulation.** We can estimate the sampling distribution of the mean of a sample of size n by drawing many samples of size n, computing the mean of each sample, and then forming a histogram of the collection of sample means. Figure 7.2 shows the results of such a simulation for sample sizes of 8, 16, 32, and 64 with 500 replications for each sample size. Three features of Figure 7.2 are noteworthy:

1. All the histograms are centered about the population mean, 814.6.
2. As the sample size increases, the histograms become less spread out.
3. Although the shape of the histogram of population values (Figure 7.1) is not symmetric about the mean, the histograms in Figure 7.2 are more nearly so.

These features will be explained quantitatively. ■

As we have said, $\overline{X}$ is a random variable whose distribution is determined by that of the X_i. We thus examine the distribution of a single sample element, X_i. It should be noted that the following lemma holds whether sampling is with or without replacement.