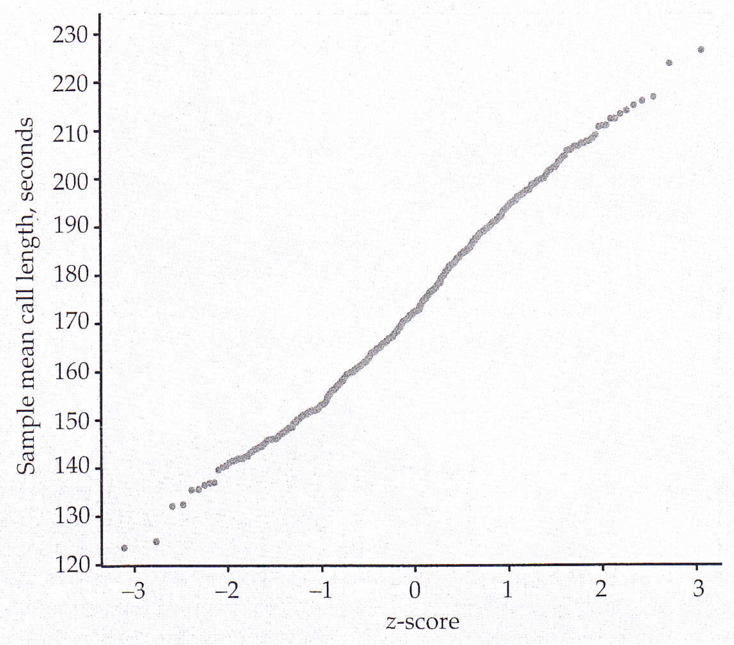

FIGURE 5.9 Normal quantile plot of the 500 sample means in Figure 5.8(b). The distribution is close to normal.

The mean and standard deviation of $\bar{x}$

The sample mean $\bar{x}$ from a sample or an experiment is an estimate of the mean μ of the underlying population, just as a sample proportion $\hat{p}$ is an estimate of a population proportion p. The sampling distribution of $\bar{x}$ is determined by the design used to produce the data, the sample size n, and the population distribution.

Select an SRS of size n from a population, and measure a variable X on each individual in the sample. The n measurements are values of n random variables $X_1, X_2, \ldots, X_n$. A single X_i is a measurement on one individual selected at random from the population and therefore has the distribution of the population. If the population is large relative to the sample, we can consider $X_1, X_2, \ldots, X_n$ to be independent random variables each having the same distribution. This is our probability model for measurements on each individual in an SRS.

The sample mean of an SRS of size n is

$$\bar{x} = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$$

If the population has mean μ, then μ is the mean of each observation X_i. Therefore, by the addition rule for means of random variables,

$$\begin{aligned}\mu_{\bar{x}} &= \frac{1}{n}(\mu_{X_1} + \mu_{X_2} + \cdots + \mu_{X_n}) \\ &= \frac{1}{n}(\mu + \mu + \cdots + \mu) = \mu\end{aligned}$$

That is, *the mean of $\bar{x}$ is the same as the mean of the population.* The sample mean $\bar{x}$ is therefore an unbiased estimator of the unknown population mean μ.

The observations are independent, so the addition rule for variances also applies:

$$\sigma^2_{\bar{x}} = \left(\frac{1}{n}\right)^2 (\sigma^2_{X_1} + \sigma^2_{X_2} + \cdots + \sigma^2_{X_n})$$
$$= \left(\frac{1}{n}\right)^2 (\sigma^2 + \sigma^2 + \cdots + \sigma^2)$$
$$= \frac{\sigma^2}{n}$$

Just as in the case of a sample proportion $\hat{p}$, the variability of the sampling distribution of a sample mean decreases as the sample size grows. Because the standard deviation of $\bar{x}$ is $\sigma/\sqrt{n}$, it is again true that the standard deviation of the statistic decreases in proportion to the square root of the sample size. Here is a summary of these facts.

MEAN AND STANDARD DEVIATION OF A SAMPLE MEAN

Let $\bar{x}$ be the mean of an SRS of size n from a population having mean μ and standard deviation σ. The mean and standard deviation of $\bar{x}$ are

$$\mu_{\bar{x}} = \mu$$
$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

σ modified by the factor $\frac{1}{\sqrt{n}}$

EXAMPLE 5.15 How accurately does a sample mean $\bar{x}$ estimate a population mean μ? Because the values of $\bar{x}$ vary from sample to sample, we must give an answer in terms of the sampling distribution. We know that $\bar{x}$ is an unbiased estimator of μ, so its values in repeated samples are not systematically too high or too low. Most samples will give an $\bar{x}$-value close to μ if the sampling distribution is concentrated close to its mean μ. So the accuracy of estimation depends on the spread of the sampling distribution.

The standard deviation of the population of service call lengths in Figure 5.8(a) is $\sigma = 184.81$ seconds. The length of a single call will often be far from the population mean. If we choose an SRS of 20 calls, the standard deviation of their mean length is

$$\sigma_{\bar{x}} = \frac{184.81}{\sqrt{20}} = 41.32 \text{ seconds}$$

Averaging over more calls reduces the variability and makes it more likely that $\bar{x}$ is close to μ. Our sample size of 80 calls is 4 times 20, so the standard deviation will be half as large:

$$\sigma_{\bar{x}} = \frac{184.81}{\sqrt{80}} = 20.66 \text{ seconds}$$