distribution to formulate ...

Standard Error of the Mean

Because a sampling distribution is a normal curve, we can also establish its variability. The standard deviation of a theoretical sampling distribution of means is called the **standard error of the mean** ($\sigma_{\overline{X}}$). This value is considered an estimate of the population standard deviation, σ. The curve in Figure 18.2A represents a hypothetical sampling distribution formed by repeated sampling of birth weights, with samples of $n = 10$. The means of such small samples tend to vary, and in fact, we see a wide curve with great variability. The sampling distribution in the curve in Figure 18.2B was constructed from the same population, but with samples of $n = 50$. These sample means form a narrower distribution curve with less variability and, therefore, a smaller standard deviation. As sample size increases, samples become more representative of the population, and their means are more likely to be closer to the population mean; that is, their sampling error will be smaller. Therefore, the standard deviation of the sampling distribution is an indicator of the degree of sampling error, reflecting how accurately the various sample means estimate the population mean.

Because we do not actually construct a sampling distribution, we need some useful way to estimate the standard error of the mean from sample data. This estimate, $s_{\overline{X}}$, is based on the standard deviation and size of the sample:

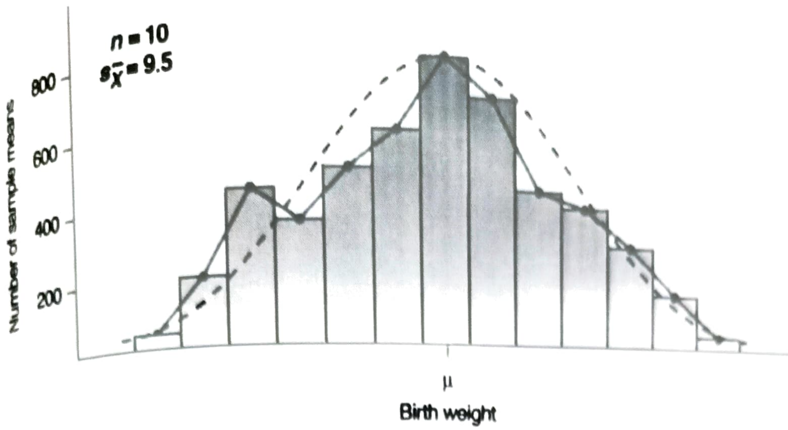
$$s_{\overline{X}} = \frac{s}{\sqrt{n}} \qquad (18.1)$$

Using our example of birth weights, for a single sample of 10 babies, we found a mean of 115 with a standard deviation of 30 (see Figure 18.2A). Therefore, $s_{\overline{X}} = 30/\sqrt{10} = 9.5$. With a sample of $n = 50$, $s_{\overline{X}} = 30/\sqrt{50} = 4.2$. As illustrated in Figure 18.2, as n increases, the standard error of the mean decreases. With larger samples the sampling distribution is expected to be less variable, and therefore, a statistic based on a large sample is considered a better estimate of a population parameter than one based on a smaller sample.

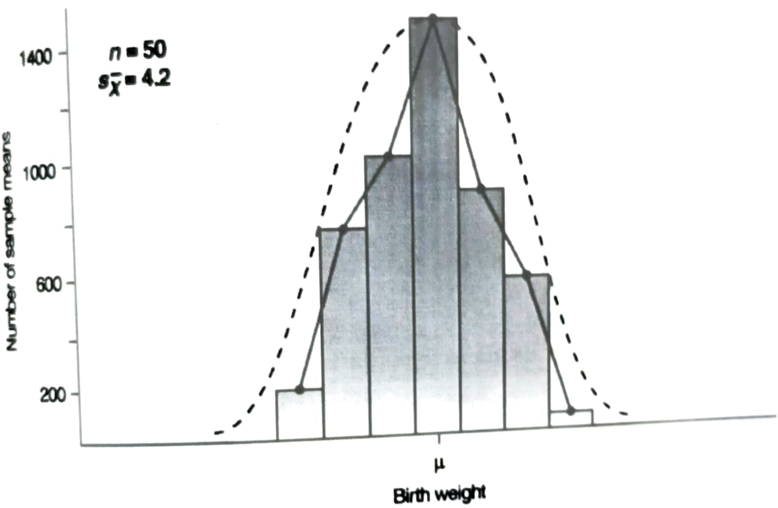
A sample mean, together with its standard error, helps us imagine what the sampling distribution curve would look like. For example, for a sample of $n = 50$, with $\overline{X} = 115$ and $s_{\overline{X}} = 4.2$, the theoretical sampling distribution might look like the curve

---

*This phenomenon is explained by the *central limit theorem*, which demonstrates that even for skewed distributions, the sampling distribution of means will approach the normal curve as n increases. Therefore, we can use sampling distributions and the probabilities associated with the normal curve to predict population characteristics for any distribution.



A



B

FIGURE 18.2 Hypothetical sampling distributions for birth weight. Curve A is drawn for samples with $n = 10$. Curve B is drawn for samples with $n = 50$.

shown in Figure 18.2B. If we use this curve as an estimate of the population distribution, we can determine the probability of drawing a single sample with a certain mean. Based on our knowledge of the normal curve, the chances are 95.45 out of 100 that any single random sample we might draw from this population would have a mean between 106.6 and 123.4 ($\pm 2s_{\bar{X}}$). Therefore, the probability is 95.45% that a sample mean will lie within this range. We can also say that there is less than a 5% chance that any sample mean drawn from this population will be less than 106.6 or above 123.4. We should note that the standard error cannot be a direct measure of variance in the population, because it is a function of sample size.

CONFIDENCE INTERVALS

For many research applications, sample data are used to estimate unknown population parameters. For example, we can sample medical records to determine length of hospital stay for patients with certain diagnoses or we could study normative values for tests of motor function. The purpose of these types of analyses is to estimate how the population behaves and to use this information for decision making or as a foundation for further research.

We can use our knowledge of sampling distributions to estimate population parameters in two ways. A **point estimate** is a single value obtained by direct calculation from sample data, such as using $\overline{X}$ to estimate μ. We know, however, that any single sample value will most likely contain some degree of error as a population estimate. Therefore, it is often more meaningful to use an **interval estimate,** by which we specify an interval within which we believe the population parameter will lie. Such an estimate takes into consideration not only the value of a single sample statistic, but the relative accuracy of that statistic as well.

For example, Fitzgerald et al[1] estimated the population mean for lumbar spinal extension for 30- to 39-year-olds. Based on a random sample of 42 individuals, they determined that $\overline{X} = 40.0$ degrees and $s = 8.8$ degrees. Therefore, the point estimate of μ is the sample mean, 40.0 degrees. How can we tell how accurate this estimate is? Perhaps we would be more comfortable giving a range of values within which we are fairly sure the population mean will fall. For instance, we might guess that the population mean is likely to be within 5 degrees of the sample mean, to fall within the interval 35 to 45 degrees. We must be more precise than guessing allows, however, in proposing such an interval, so that we can be "confident" that the interval is an accurate estimate.

A **confidence interval (CI)** is a range of scores with specific boundaries, or *confidence limits*, that should contain the population mean. The boundaries of the confidence interval are based on the sample mean and its standard error. The wider the interval we propose, the more confident we will be that the true population mean will fall within it. This degree of confidence is expressed as a probability percentage, such as 95% confidence.

To illustrate the procedure for constructing a 95% confidence interval, consider the example of lumbar spine extension, with $\overline{X} = 40.0$, $s = 8.8$, $n = 42$, and $s_{\overline{X}} = 8.8/\sqrt{42} = 1.36$. The sampling distribution estimated from this sample is shown in Figure 18.3. We know that 95.45% of the total distribution will fall within $\pm 2s_{\overline{X}}$ from the mean, or within the boundaries of $z = \pm 2$. Therefore, to determine the proportion of the curve within 95%, we need to determine points just slightly less than $z = \pm 2$. By referring to Table A.1 in the Appendix, we can determine that 0.95 of the total curve (0.475 on either side of the mean) is bounded by a z-score of ± 1.96, just less than 2 standard error units above and below the mean. Therefore, as shown in Figure 18.3, 95% of the total sampling distribution will fall between $-1.96s_{\overline{X}}$ and $+1.96s_{\overline{X}}$. We are 95% sure that the population mean will fall within this interval. This is called the *95% confidence interval.*

We obtain the boundaries of a confidence interval using the formula

$$CI = \overline{X} \pm (z)s_{\overline{X}} \tag{18.2}$$

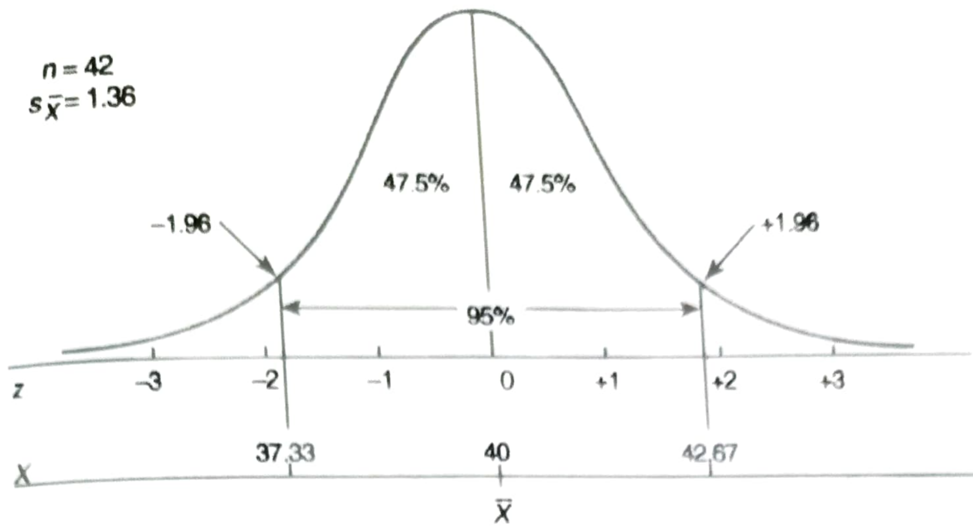


FIGURE 18.3 95% Confidence interval for sampling distribution of lumbar extension range of motion for 30–39 year olds.

For 95% confidence intervals, z = ±1.96.

For our data, therefore,

$$95\%\ \text{CI} = 40.0 \pm (1.96)(1.36)$$

$$= 40.0 \pm 2.67$$

$$95\%\ \text{CI} = 37.33, 42.67$$

We are 95% confident that the population mean of lumbar extension for 30 to 39-year-olds will fall between 37.33 and 42.67 degrees.

How can we interpret this statement? Because of sampling error, one sample we select may have a mean of 50 degrees, with 95% confidence limits between 40 and 60 degrees. Another sample could have a mean of 52 degrees, with 95% confidence limits between 42 and 62 degrees. The 95% confidence limits indicate that if we were to draw 100 random samples, each with $n = 42$, we could construct 100 confidence intervals around the sample means, 95 of which could be expected to contain the true population mean, as illustrated in Figure 18.4. Five of the 100 intervals would not contain the population mean. This would occur just by chance, because the scores chosen for those five samples would be too extreme and not good representatives of the population. In reality, however, we construct only one confidence interval based on the data from only one sample. Theoretically, then, we cannot know if that one sample would produce one of the 95 correct intervals or one of the 5 incorrect ones. Therefore, there is a 5% chance that the population mean is not included in the obtained interval, that is, a 5% chance the interval is one of the incorrect ones.

To be more confident of the accuracy of an interval, we could construct a 99% confidence interval, allowing only a 1% risk that the interval we propose will not contain