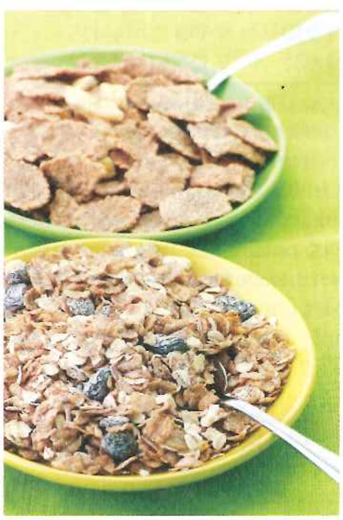
a. Describe the population from which the random sample was drawn.
b. Use a 98% confidence interval to estimate the mean revenue of the population of companies in question.
c. Interpret your confidence interval in the context of the problem.
d. What characteristic must the population possess to ensure the appropriateness of the estimation procedure used in part **b?**
e. Suppose *Forbes* reports that the true mean revenue of the 212 companies on the list is \$5.0 billion. Is the claim believable?

6.4 Large-Sample Confidence Interval for a Population Proportion

The number of public opinion polls has grown at an astounding rate in recent years. Almost daily, the news media report the results of some poll. Pollsters regularly determine the percentage of people who approve of the president's on-the-job performance, the fraction of voters in favor of a certain candidate, the fraction of customers who prefer a particular product, and the proportion of households that watch a particular TV program. In each case, we are interested in estimating the percentage (or proportion) of some group with a certain characteristic. In this section, we consider methods for making inferences about population proportions when the sample is large.

Example 6.6

Estimating a Population Proportion—Preference for Breakfast Cereal



Problem A food-products company conducted a market study by randomly sampling and interviewing 1,000 consumers to determine which brand of breakfast cereal they prefer. Suppose 313 consumers were found to prefer the company's brand. How would you estimate the true fraction of *all* consumers who prefer the company's cereal brand?

Solution In this study, consumers are asked which brand of breakfast cereal they prefer. Note that "brand" is a qualitative variable and that what we are asking is how you would estimate the probability p of success in a binomial experiment, where p is the probability that a chosen consumer prefers the company's brand. One logical method of estimating p for the population is to use the proportion of successes in the sample—that is, we can estimate p by calculating

$$\hat{p} = \frac{\text{Number of consumers sampled who prefer the company's brand}}{\text{Number of consumers sampled}}$$

where $\hat{p}$ is read "p hat." Thus, in this case,

$$\hat{p} = \frac{313}{1,000} = .313$$

Look Back To determine the reliability of the estimator $\hat{p}$, we need to know its sampling distribution—that is, if we were to draw samples of 1,000 consumers over and over again, each time calculating a new estimate $\hat{p}$, what would be the frequency distribution of all the $\hat{p}$ values? Recall (Section 5.4) that the answer lies in viewing $\hat{p}$ as the average, or mean, number of successes per trial over the n trials. If each success is assigned a value equal to 1 and a failure is assigned a value of 0, then the sum of all n sample observations is x, the total number of successes, and $\hat{p} = x/n$ is the average, or mean, number of successes per trial in the n trials. The Central Limit Theorem tells us that the *relative frequency distribution of the sample mean for any population is approximately normal for sufficiently large samples.*

■ Now Work Exercise 6.49a

The repeated sampling distribution of $\hat{p}$ was the topic of Section 5.4. The characteristics are repeated in the next box and shown in Figure 6.12.

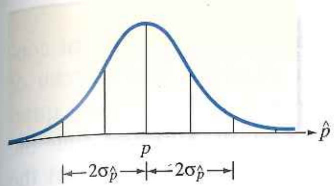


Figure 6.12
Sampling distribution of $\hat{p}$

Properties of the Sampling Distribution of $\hat{p}$

1. The mean of the sampling distribution of $\hat{p}$ is p; that is, $\hat{p}$ is an unbiased estimator of p.
2. The standard deviation of the sampling distribution of $\hat{p}$ is $\sqrt{pq/n}$; that is, $\sigma_{\hat{p}} = \sqrt{pq/n}$, where $q = 1 - p$.
3. For large samples, the sampling distribution of $\hat{p}$ is approximately normal. A sample size is considered large if both $n\hat{p} \geq 15$ and $n\hat{q} \geq 15$.

Note: our textbook (Devore) uses 10, not 15 here.

The fact that $\hat{p}$ is a "sample mean number of successes per trial" allows us to form confidence intervals about p in a manner that is completely analogous to that used for large-sample estimation of μ.

Large-Sample Confidence Interval for $\hat{p}$

$$\hat{p} \pm z_{\alpha/2}\sigma_{\hat{p}} = \hat{p} \pm z_{\alpha/2}\sqrt{\frac{pq}{n}} \approx \hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

where $\hat{p} = \frac{x}{n}$ and $\hat{q} = 1 - \hat{p}$

Note: When n is large, $\hat{p}$ can approximate the value of p in the formula for $\sigma_{\hat{p}}$.

Conditions Required for a Valid Large-Sample Confidence Interval for p

1. A random sample is selected from the target population.
2. The sample size n is large. (This condition will be satisfied if both $n\hat{p} \geq 15$ and $n\hat{q} \geq 15$. Note that $n\hat{p}$ and $n\hat{q}$ are simply the number of successes and number of failures, respectively, in the sample.)

Thus, if 313 of 1,000 consumers prefer the company's cereal brand, a 95% confidence interval for the proportion of *all* consumers who prefer the company's brand is

$$\hat{p} \pm z_{\alpha/2}\sigma_{\hat{p}} = .313 \pm 1.96\sqrt{\frac{pq}{1{,}000}}$$

where $q = 1 - p$. Just as we needed an approximation for σ in calculating a large-sample confidence interval for μ, we now need an approximation for p. As Table 6.6 shows, the approximation for p does not have to be especially accurate because the value of $\sqrt{pq}$ needed for the confidence interval is relatively insensitive to changes in p. Therefore, we can use $\hat{p}$ to approximate p. Keeping in mind that $\hat{q} = 1 - \hat{p}$, we substitute these values into the formula for the confidence interval:

$$\hat{p} \pm 1.96\sqrt{\frac{pq}{1{,}000}} \approx \hat{p} \pm 1.96\sqrt{\frac{\hat{p}\hat{q}}{1{,}000}}$$
$$= .313 \pm 1.96\sqrt{\frac{(.313)(.687)}{1{,}000}}$$
$$= .313 \pm .029$$
$$= (.284, .342)$$

Table 6.6 Values of pq for Several Different Values of p

p	pq	$\sqrt{pq}$
.5	.25	.50
.6 or .4	.24	.49
.7 or .3	.21	.46
.8 or .2	.16	.40
.9 or .1	.09	.30

The company can be 95% confident that the interval from 28.4% to 34.2% contains the true percentage of *all* consumers who prefer its brand—that is, in repeated construction of confidence intervals, approximately 95% of all samples would produce confidence intervals that enclose p. Note that the guidelines for interpreting a confidence interval about μ also apply to interpreting a confidence interval for p because p is the "population fraction of successes" in a binomial experiment.