Example 2 Confidence Interval for Claim About Denominations

Use the sample data given in Example 1 to construct a 90% confidence interval estimate of the difference between the two population proportions. (As shown in Table 8-1 in Section 8-2, a one-tailed hypothesis test with significance level $\alpha = 0.05$ requires a confidence level of 90%.) What does the result suggest about the claim that "money in a large denomination is less likely to be spent relative to an equivalent amount in many smaller denominations"?

Solution

Requirement check We are using the same data from Example 1, and the same requirement check applies here, so the requirements are satisfied. ✔

With a 90% confidence level, $z_{\alpha/2} = 1.645$ (from Table A-2). We first calculate the value of the margin of error E as shown here.

$$E = z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}} = 1.645\sqrt{\frac{\left(\frac{12}{46}\right)\left(\frac{34}{46}\right)}{46} + \frac{\left(\frac{27}{43}\right)\left(\frac{16}{43}\right)}{43}}$$
$$= 0.161387$$

With $\hat{p}_1 = 12/46 = 0.260870$ and $\hat{p}_2 = 27/43 = 0.627907$, $\hat{p}_1 - \hat{p}_2 = -0.367037$. With $\hat{p}_1 - \hat{p}_2 = -0.367037$ and $E = 0.161387$, the confidence interval is evaluated as follows, with the confidence interval limits rounded to three significant digits:

$$(\hat{p}_1 - \hat{p}_2) - E < (p_1 - p_2) < (\hat{p}_1 - \hat{p}_2) + E$$
$$-0.367037 - 0.161387 < (p_1 - p_2) < -0.367037 + 0.161387$$
$$-0.528 < (p_1 - p_2) < -0.206$$

Interpretation

The confidence interval limits do not contain 0, implying that there is a significant difference between the two proportions. The confidence interval suggests that the value of p_1 is less than the value of p_2, so there does appear to be sufficient evidence to support the claim that "money in a large denomination is less likely to be spent relative to an equivalent amount in many smaller denominations."

Rationale: Why Do the Procedures of This Section Work? The test statistic given for hypothesis tests is justified by the following reasoning.

With $n_1\hat{p}_1 \geq 5$ and $n_1\hat{q}_1 \geq 5$, the distribution of $\hat{p}_1$ can be approximated by a normal distribution with mean p_1, standard deviation $\sqrt{p_1q_1/n_1}$, and variance p_1q_1/n_1 (based on Sections 6-7 and 7-2). They also apply to the second sample. Because the distributions of $\hat{p}_1$ and $\hat{p}_2$ are each approximated by a normal distribution, the difference $\hat{p}_1 - \hat{p}_2$ will also be approximated by a normal distribution with mean $p_1 - p_2$ and variance

$$\sigma^2_{(\hat{p}_1-\hat{p}_2)} = \sigma^2_{\hat{p}_1} + \sigma^2_{\hat{p}_2} = \frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}$$

(The result above is based on this property: The variance of the *differences* between two independent random variables is the *sum* of their individual variances.) The pooled estimate of the common value of p_1 and p_2 is $\bar{p} = (x_1 + x_2)/(n_1 + n_2)$. If we

New Technology, New Data, New Insight

Residents of New York City believed that taxi cabs became scarce around rush hour in the late afternoon. Complaints could not be addressed, because there were no data to support that alleged shortage. However, as GPS units were installed on cabs, officials became able to track their locations. It was found that 20% fewer cabs were in service between 4:00 PM and 5:00 PM than in the preceding hour. Two factors were found to be responsible for this decrease. First, the 12-hour shifts were scheduled to change at 5:00 PM so that drivers on both shifts would get an equal share at a rush hour. Second, rising rents in Manhattan forced many cab companies to house their cabs in Queens, so drivers had to start returning around 4:00 PM so that they could make it back in time and avoid fines for being late. As of this writing, officials are considering regulations designed to eliminate or reduce the late-afternoon shortage. Any changes will result from planners' turning to the new GPS technology for data that could be objectively analyzed, instead of relying on subjective beliefs and anecdotal stories.