OPINION POLLS: USING THE CONFIDENCE INTERVAL FOR A PROPORTION

We saw in the textbook near the end of Chapter 8 that Margie Novera, the news anchorperson on Channel 3, said about the election: “A recent poll of 250 voters showed that candidate Jones has 64 percent of the votes. The margin of error is plus or minus 6 percent.” We observed that “margin of error” is Margie’s way of referring to a 95% confidence interval for a population proportion, which is simply a special kind of population mean. Here we derive the pollster’s confidence interval that we simply reported in the textbook.

Let the variable $X$ have only two possible values: $X_i = 1$ if the $i$th respondent says, “Yes, I favor candidate Jones,” and $X_i = 0$ if the respondent says, “No, I do not favor candidate Jones.” Then $\sum X_i$ is the total number of respondents who favor Jones. In our example, $\sum X_i = .64(250) = 160$ respondents who favored Jones.

We need some notation for proportions. We call $p$ the proportion of respondents who say yes; then $1 - p$ is the proportion who say no. Thus, $p = \sum X_i/n = 160/250 = .64$, and $1 - p = 1 - .64 = .36$. Note that by its definition $p = \bar{X}$. The population proportion (whose value we would like to know) is the parameter we call $\pi$ (Greek capital “pi”), and $\pi$ is simply a special name for $\mu$ when we are discussing proportions. Thus, $p$ (another name for $\bar{X}$) is the best point-estimator for $\pi$ (another name for $\mu$). Recall that the confidence interval for the mean when $\sigma$ is unknown is Equation (8.3):

$$\bar{X} - t_{cv}(s_{\bar{X}}) \leq \mu \leq \bar{X} + t_{cv}(s_{\bar{X}})$$

We just saw that $\bar{X} = p$. What about $s_{\bar{X}}$? We know that $s_{\bar{X}} = \frac{s}{\sqrt{n}}$, and we also know that

$$s = \sqrt{\frac{\sum (X_i - \bar{X})^2}{n - 1}}.$$

Because in this case $X_i$ can take values of only 0 or 1, we see that $X_i^2 = X_i$. That fact (and a little algebra) allows us to show that the numerator of $s$ is $np(1 - p)$:

$$\sum (X_i - \bar{X})^2 = \sum (X_i^2 - 2X_i\bar{X} + \bar{X}^2) = \sum X_i^2 - 2\bar{X}\sum X_i + \sum \bar{X}^2 = np - 2p(np) + np^2 = np(1 - p)$$

Then $s = \sqrt{\frac{\sum (X_i - \bar{X})^2}{n - 1}} = \sqrt{\frac{np(1 - p)}{n - 1}}$, so $s_{\bar{X}} = \frac{s}{\sqrt{n}} = \sqrt{\frac{np(1 - p)}{n(n - 1)}} = \frac{\sqrt{p(1 - p)}}{(n - 1)}$. 

---

The opinion poll result is just a confidence interval if we substitute $p$ for $\bar{X}$ and $\pi$ for $\mu$. 
In the proportion situation, where $\bar{X} = p$, we see that $s_{\bar{X}} = s_{p}$, which we call the *standard error of a proportion*. Therefore

$$s_{p} = \sqrt{\frac{p(1-p)}{n-1}}$$

$$= \sqrt{\frac{.64(1-.64)}{249}} = .03$$

The confidence interval is then defined exactly as in Equation (8.3):

$$p - t_{cv}(s_{p}) \leq \Pi \leq p + t_{cv}(s_{p})$$

or, because $p = .64$ and $df = n - 1 = 249$,

$$.64 - 1.96(.03) \leq \Pi \leq .64 + 1.96(.03)$$

or approximately

$$.58 \leq \Pi \leq 1.70$$

Thus, we can say with 95% confidence that the proportion of the voter population who favor Jones is 64% ± 6%, which is exactly what Margie reported.

Pollsters often ask several questions in the same poll, and rather than compute a separate standard error of a proportion for each question, they use the same “largest possible” standard error for all the questions. The standard error is the largest when $p = .5$, in which case $s_{p}$ becomes

$$s_{p} = \sqrt{\frac{p(1-p)}{n-1}} = \sqrt{\frac{.5(1-.5)}{n-1}} = \sqrt{\frac{.25}{n-1}}$$