relative to the size N of the population (more than about 10% of the population size), then you would use a correction factor for the formula for the margin of error that makes it smaller. Specifically, an approximate confidence interval is

$$\hat{p} = z^* \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n-1}}\sqrt{\frac{N-n}{N-1}}$$

where N is the size of the population.

AP Sample Test

AP1. E. When the sample is selected at random, the P-value is the probability of getting a sample proportion as extreme as or more extreme than that actually observed, given that null hypothesis is true. In this case, Ms. Chang can only ask whether the proportion of her students who get the question correct is consistent with what would be expected from a random sample of students nationwide.

AP2. B. The margin of error is inversely proportional to the square root of n, so if four times as many people were surveyed, the margin of error would be half as large.

AP3. C. The 95% confidence interval is (0.45579, 0.59421). This means that it is plausible that the percentage of all alumni who favor abolishing the school dress code is 50% or less.

AP4. A. Because 0 is in the confidence interval, it is plausible that there is no difference between the two teachers. Choice E is correct because although the difference in pass rates was 20%, that is not statistically significant because of the width of the confidence interval, which is a result of sample sizes of only 25 students each.

AP5. C. The probability that the null hypothesis will be rejected when it is true is equal to α.

AP6. B. All conditions are satisfied for a two-sample test for the difference of two proportions. The z-score is about -1.92 with a one-sided P-value of 0.027. Therefore, reject the null hypothesis that there is no difference between the proportion of yellow M&M's and the proportion of yellow Skittles. The difference in the two proportions cannot reasonably be attributed to chance variation alone.

AP7. C. Choice E is close, but technically incorrect. See page 479 in the student book.

AP8. E

AP9. **a.** 45

b. Forty intervals, or 80%, actually captured the population proportion of 0.20.

c. Most of the intervals are too close to 0 and shorter than average. A sample proportion that underestimates the value of p will also underestimate the standard error, and the resulting interval will be too close to 0 and too short. (Similarly, suppose the true p is greater than 0.5. A sample proportion that overestimates this value of p will also underestimate the standard error, and the resulting interval will be too close to 1.00 and too short.)

d. You are using a continuous distribution to approximate a discrete distribution. Students constructed the horizontal lines in Display 8.2 using the fact that 95% of all values in a normal distribution lie within 1.96 standard deviations of the mean; that is, the middle 95% of a normal distribution was used as the middle 95% of the binomial distribution centered in the same spot. For most binomial distributions, it's not possible to get an exact 95% middle. Further, all binomial distributions, except those where $p = 0.5$, are skewed. The normal approximation is symmetric. Finally, the sample proportion was used to estimate the population proportion in the formula for the standard error. Thus, two quantities are being estimated (center and spread) by the same sample proportion. All of these approximations can cause the capture rate to differ from its nominal value.

AP10. **a.** The standard 95% confidence interval is

$$\hat{p} \pm 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.1 \pm 1.96\sqrt{\frac{0.1 \cdot 0.9}{50}}$$

or about 0.017 to 0.183.

The Plus 4 confidence interval using $\tilde{p} = \frac{x+2}{n+4} = 754 \approx 0.130$ is

$$\tilde{p} \pm 1.96\sqrt{\frac{\tilde{p}(1-\tilde{p})}{n+4}} = 0.130 \pm 1.96\sqrt{\frac{0.130 \cdot 0.870}{54}}$$

or about 0.040 to 0.220.

The center changed from 0.10 to 0.13, which is a move toward 0.5. The length of the interval increased from about 0.166 to 0.180.

b. The standard interval is $0.4 \pm 1.96\sqrt{\frac{0.4 \cdot 0.6}{50}}$, or about 0.264 to 0.536.

The Plus 4 interval is $0.407 \pm 1.96\sqrt{\frac{0.407 \cdot 0.593}{54}}$, or about 0.276 to 0.538.

The center moved toward 0.5 (from 0.4 to 0.407), but the length of the interval decreased from 0.272 to 0.262. Both changes were smaller for $x = 20$ than for $x = 5$. The impact was greater on the interval based on the smaller sample proportion.

c. Forty-five of the 50 intervals, or 90%, captured the population proportion.

d. The intervals that do not capture the true value of p are more balanced on either side of p, the intervals themselves are of more uniform length, and the capture rate is about what you would expect for 90% confidence intervals. Specifically, when $\hat{p}$ is close to 0 (or close to 1), the Plus 4 interval tends to be longer than the standard interval. When $\hat{p}$ is close to 0.5, the Plus 4 interval tends to be about the same length as the standard interval.

Notes on the Plus 4 Confidence Interval

Terminology: The applet at www.rossmanchance.com/applets demonstrates the comparison between the standard and Plus 4 intervals. The Plus 4 method is sometimes called the Wilson confidence internal or adjusted Wald confidence interval.

Justification of the Interval: One way to justify the Plus 4 method of adjusting confidence intervals is to work backward from a test of significance. The standard test statistic for testing the hypothesis that $p = p_0$ is

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}}$$

For a level α two-tailed test, the null hypothesis is *not* rejected if

$$\left| \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}} \right| \le z^*$$

where z^* is the value that cuts off an area of $\frac{\alpha}{2}$ in the upper tail of a standard normal distribution. Any value of p_0 for which the inequality holds will not be rejected by the test. In other words, any value of p_0 that satisfies the inequality is a plausible value for p. The set of these plausible values is a confidence interval for p.

Finding these plausible values merely takes a little algebra. Squaring both sides of the inequality and multiplying both sides by the denominator gives

$$(\hat{p} - p_0)^2 \le \frac{p_0(1 - p_0)}{n}(z^*)^2$$

To find the endpoints of the confidence interval, solve $(\hat{p} - p_0)^2 = \frac{p_0(1 - p_0)}{n}(z^*)^2$ for p_0. This is a quadratic equation in p_0 and so has two solutions, with the values that satisfy the inequality lying between them. These endpoints are

$$\frac{\hat{p} + \frac{(z^*)^2}{2n} \pm z^* \sqrt{\frac{\hat{p}(1 - \hat{p}) + \frac{(z^*)^2}{4n}}{n}}}{1 + \frac{(z^*)^2}{n}}$$

or

$$\hat{p}\left(\frac{n}{n + (z^*)^2}\right) + \frac{1}{2}\left(\frac{(z^*)^2}{n + (z^*)^2}\right) \pm z^* \frac{\sqrt{\frac{4n\hat{p}(1 - \hat{p}) + (z^*)^2}{4n^2}}}{\left(\frac{1 + (z^*)^2}{n}\right)}$$

This looks messy, but note that the midpoint of the interval, $\hat{p}\left(\frac{n}{n + (z^*)^2}\right) + \frac{1}{2}\left(\frac{(z^*)^2}{n + (z^*)^2}\right)$ is a weighted average between $\hat{p}$ and $\frac{1}{2}$. Observe that as n gets large, the weighted average converges to the original $\hat{p}$.

For a 95% confidence interval z^* is approximately 2, and the midpoint of the confidence interval simplifies to

$$\begin{aligned}\hat{p}\left(\frac{n}{n + (z^*)^2}\right) + \frac{1}{2}\left(\frac{(z^*)^2}{n + (z^*)^2}\right) &= \hat{p}\left(\frac{n}{n + 4}\right) + \frac{1}{2}\left(\frac{4}{n + 4}\right) \\ &= \frac{x}{n}\left(\frac{n}{n + 4}\right) + \frac{1}{2}\left(\frac{4}{n + 4}\right) \\ &= \frac{x + 2}{n + 4} \\ &= \tilde{p}\end{aligned}$$

where x denotes the number of successes in the sample. The adjusted margin of error is approximated by using the standard formula with $\tilde{p}$ replacing $\hat{p}$ and $n + 4$ replacing n as the sample size. In final form, the Plus 4 confidence interval is simply

$$\tilde{p} \pm z^* \sqrt{\frac{\tilde{p}(1 - \tilde{p})}{n + 4}}$$

Recall that we substituted 2 for z^*, so this adjusted interval works best for a 95% confidence interval. But, it works well for any level of confidence between 90% and 99%.

Additional Exercises

If your class is particularly mathematically inclined, you may wish to discuss the following questions:

- Show, by numerical arguments or algebra, that $\tilde{p}$ is always between $\hat{p}$ and 0.5.
- What happens to the relationship between $\hat{p}$ and $\tilde{p}$ as n gets large?