# Solutions

## In-Class Activities

### Activity 16-1: Generation M

**a.** The observational units are American youth ages 8–18.

**b.** The variable is *whether each youth has a television in his or her bedroom* (binary categorical).

**c.** The value .68 is a statistic, which can be denoted by the symbol $\hat{p}$.

**d.** The relevant parameter of interest is the population proportion of all American youth ages 8–18 who have a television in their bedrooms, which can be denoted by the symbol $\pi$.

**e.** The Kaiser survey does not allow the researchers to determine the exact value of the parameter because they did not survey all American youth in the population.

**f.** The parameter value is more likely to be close to the survey's sample proportion than to be far from it, because the surveyed sample was randomly selected. You expect some sampling variability, but with a large sample size like this, you expect the sample proportion to be reasonably close to the true parameter.

**g.** The value $\hat{p}$ (.68) seems like a reasonable replacement to use as an estimate for $\pi$.

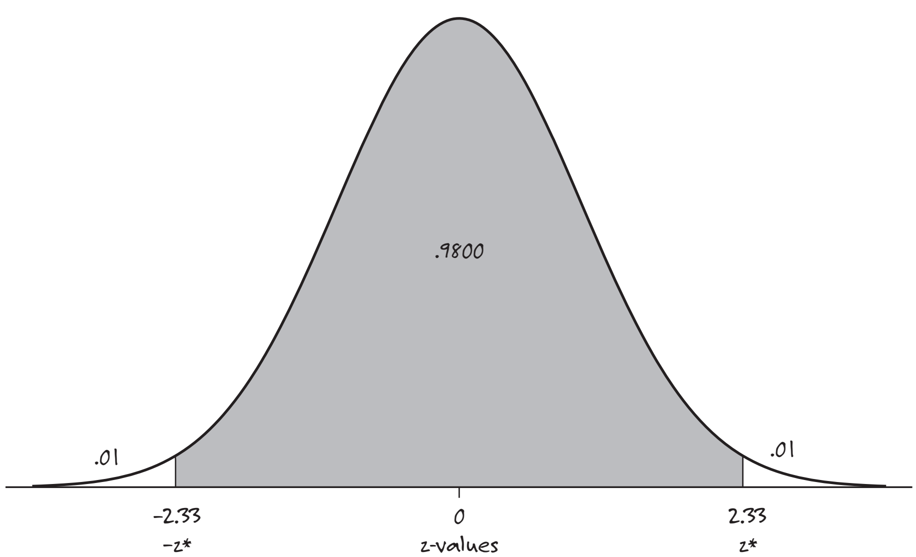
**h.** The standard error of $\hat{p}$ is $\sqrt{\frac{(.68)(.32)}{2032}} = .01035$.

**i.** You calculate $.68 \pm 2(.01035) = .68 \pm .0207 = (.6593, .7007)$. Consider the value of the population parameter, $\pi$, to be somewhere in this interval.

**j.** You do not know for sure whether the actual value of $\pi$ is contained in this interval (as in part e).

## Activity 16-2: Critical Values

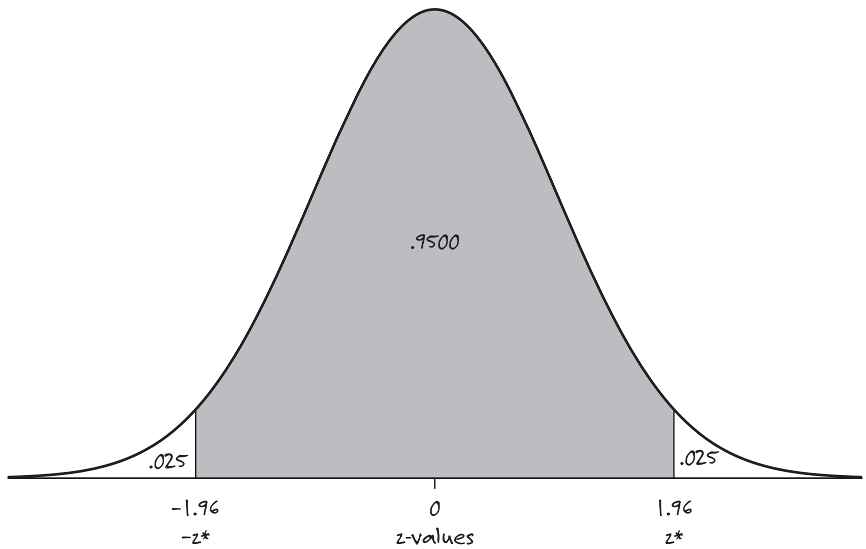
**a.** Here is the sketch:




**b.** The total area to the left of $z^*$ is .9900.

**c.** $z^* = 2.33$

**d.** Here is the sketch:




The total area to the left of $z^*$ is .975.

$z^* = 1.96$



## Activity 16-3: Generation M

**a.** For a 95% confidence interval, you calculate $\sqrt{.68(.32)/2032} = .01035$. So $.68 \pm 1.96(.01035) = .68 \pm .020286 = (.6597, .7003)$.

**b.** You are 95% confident the population proportion of all American 8–18-year-olds who have a television in their bedrooms is between .66 and 70.

**c.** No, you cannot be *certain* that this interval contains the actual value of $\pi$.

**d.** Width = .7003 − .6587 = .0406

**e.** Half-width = .0406/2 = .0203

**f.** This half-width is also $z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ or 1.96 × (.01035), which is the margin-of-error.

**g.** Midpoint = (.7003 + .6597)/2 = .68

**h.** Yes, this value looks familiar. This value is $\hat{p}$. This value makes sense because the interval is created extending an equal distance above and below the observed sample proportion $\hat{p}$.

**i.** Technical conditions: (i) The sample is a simple random sample from the population of interest (probably more complicated than a "simple random sample," but still random), and (ii) the sample size is large relative to $\pi$: $(2032)(.68) = 1382 > 10$ and $2032(.32) = 650 > 10$.

**j.** Answers will vary based on student intuition, but students should expect the interval to be wider; if they specify more values, they will be more confident that the actual value is in the specified range.

**k.** You calculate $.68 \pm 2.576(.01035) = .68 \pm .0267 = (.653, .707)$.

**l.** The midpoints of both intervals are the same (.68), but the 99% confidence interval is wider; it has a greater margin-of-error (.267) vs. (.203).

**m.** .50: not plausible .75: not plausible Two-thirds: plausible

Explanation: Both .50 and .75 do not seem to be plausible values for $\pi$ as they are not contained in either the 95% or 99% confidence intervals. However, .67 is contained in both confidence intervals, so it seems to be a plausible value for $\pi$.

**n.** Boys: $.72 \pm (1.96)\sqrt{.72(.28)/996} = .72 \pm (1.96)(.0142) = (.692, .748)$

Girls: $.64 \pm (1.96)\sqrt{.64(.36)/1036} = .64 \pm (1.96)(.0149) = (.611, .669)$

**o.** These intervals do seem to indicate that there is a difference in the population proportion of boys and girls who have a television in their bedrooms. You are 95% confident the population proportion of boys with a television in their bedrooms is at least .69 (and no more than .75), whereas you are 95% confident that the proportion of girls with a television in their bedrooms is between .61 and .67. There is no overlap in these intervals; the values of $\pi$ that are plausible values for boys are not plausible values for girls.

**p.** The margins-of-error for these intervals are .028 (boys) and .029 (girls). These are greater than the margin-of-error based on the entire sample, which makes sense because the entire sample is roughly twice as large as the single gender samples, so you would expect it to have less sampling variability and therefore a smaller margin-of-error.

**q.** You set the margin-of-error formula equal to .01 and then solve for the sample size $n$, as follows:

$$.01 = 1.96\sqrt{\frac{(.68)(.32)}{n}} \qquad n = (.68)(.32)\left(\frac{1.96}{.01}\right)^2 = 8359.322;\ n = 8{,}360$$

**r.** Answers will vary by student expectation, but the required sample size will increase.

**s.** To determine the sample size, you calculate

$$.01 = 2.576\sqrt{\frac{(.68)(.32)}{n}} \qquad n = (.68)(.32)\left(\frac{2.576}{.01}\right)^2 = 14439.448;\ n = 14{,}440$$

## Activity 16-4: Candy Colors

Answers will vary. Here is one representative set of answers.

**a.** Endpoints: .315, .539.

Yes, this interval succeeds in capturing $\pi = .45$.

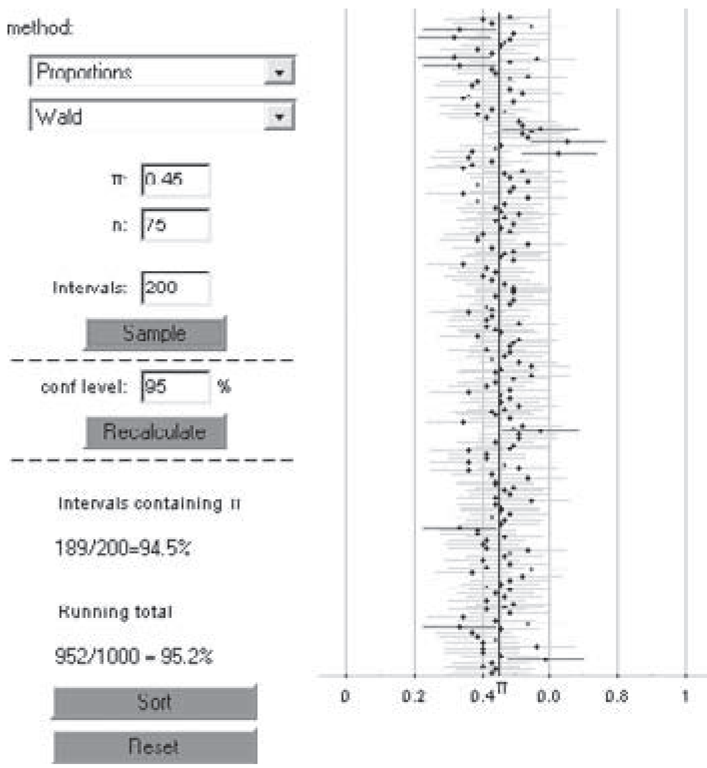
**b.** Yes, the interval changed each time. No, the intervals did not all succeed in capturing the true value of $\pi$ (4 out of 5 did).

**c.** In this simulation, the proportion of intervals that succeed in capturing $\pi$ is 190/200 = 95% (you expect something close to 95% but it may not match exactly with 200 samples).

**d.** The intervals that fail to capture $\pi$ have midpoints that are fairly far away (more than 2 standard deviations) from .45 (either on the low end of the scale or the high end).

**e.** No; if you had taken a single sample in a real situation, you would have no way of knowing whether the true value of $\pi$ was contained in your interval because you would not know what $\pi$ was; you are not guaranteed that the constructed interval will capture the value of $\pi$.

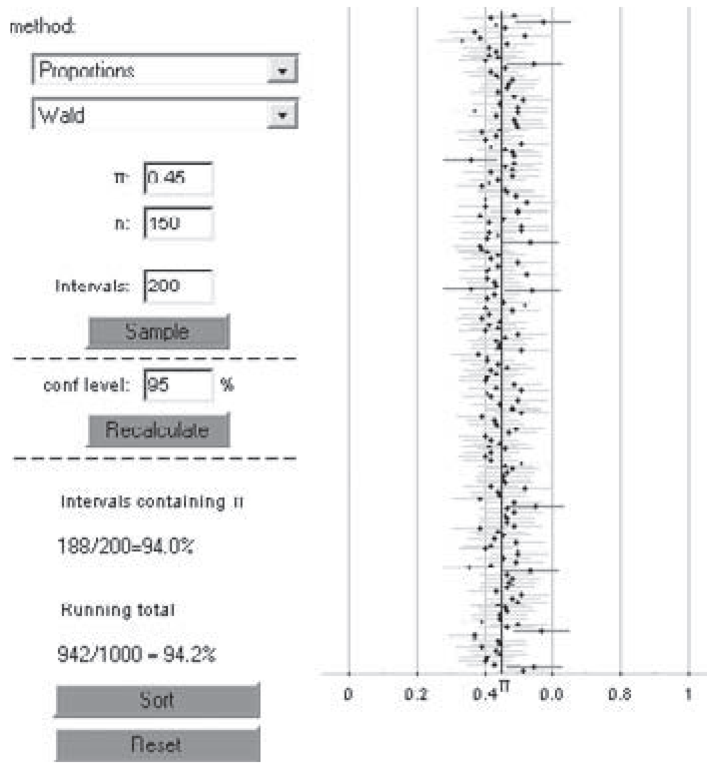
**f.** In the following simulation, the proportion of intervals that succeed in capturing $\pi$ is 952/1000 = 95.2%:




**g.** Yes, this percentage is close to 95%. Yes, it should be close to 95% because these are 95% confidence intervals; the procedure should be "successful" 95% of the time. This simulation reveals that the phrase "95% confidence" indicates that your method of creating confidence intervals is successful in capturing the true population parameter ($\pi$) 95% of the time and that it fails 5% of the time in the long run (over many, many intervals).

**h.** Answers will vary by student prediction, but the intervals will become less wide (more narrow) as the sample size increases.

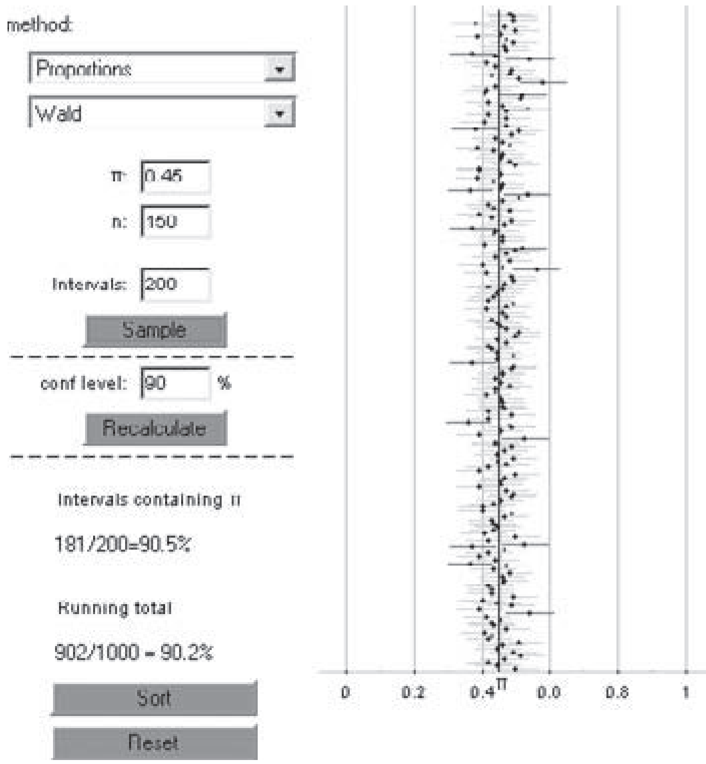
**i.** In the following simulation, the proportion of intervals that succeed in capturing $\pi$ is 942/1000 = 94.2%:




This is reasonably close to the percentage from part f. The noticeable difference about these intervals is that they are not as wide as those that were generated with samples of 75 candies. [Example interval: (.341, .499).]

**j.** Answers will vary by student prediction, but students should predict that the length of the intervals will *shorten* and the success rate will *decrease* when the confidence level is changed to 90%.

**k.** In the following simulation, the proportion of intervals that succeed in capturing $\pi$ is 902/1000 = 90.2%.




This is not particularly close to the percentages from parts f and i, but it shouldn't be because you changed the confidence level. The noticeable differences about these intervals are that they are not as wide as the previous intervals [example interval: (.334, .466)], nor are they as successful in capturing $\pi$.

## Activity 16-5: Elvis Presley and Alf Landon

**a.** For a 99.9% confidence interval, you calculate $.57 \pm (3.291)\sqrt{.57(.43)/2400000} = .57 \pm (3.22291)(.00032) = (.569, .571)$.

**b.** This interval is so narrow because the sample size is so very, very large.

**c.** The confidence interval did such a poor job predicting the election results because the necessary technical conditions were not satisfied. In particular, the sample was not randomly selected; it was chosen in a very biased fashion and therefore vastly overestimated the support for the Republican candidate.

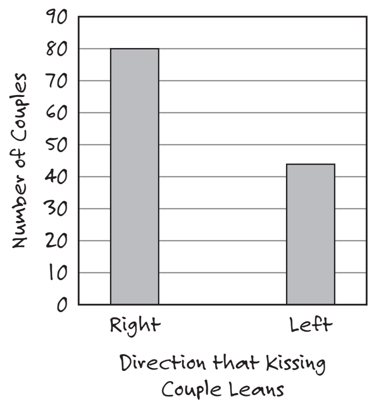
## Activity 16-6: Kissing Couples

**a.** The observational units are the kissing couples. The variable is *which direction the couples lean their heads while kissing.*

**b.** The sample consists of the 124 kissing couples observed by the researchers in various public places. The statistic is the sample proportion of couples who lean to the right when kissing:

$\hat{p} = 80/124 = .645$

A bar graph is shown here:




**c.** A 95% confidence interval for the population proportion of all couples who lean to the right is

$$.645 \pm 1.96\sqrt{\frac{(.645)(1-.645)}{124}}$$

which is .645 ± .084, or (.561, .729). You are 95% confident that the population proportion of all kissing couples who lean to the right is somewhere between .561 and .729. This "95% confidence" means that if you were to take many random samples and generate a 95% confidence interval (CI) from each, then in the long run, 95% of the resulting intervals would succeed in capturing the actual value of the population proportion, in this case the proportion of all kissing couples who lean their heads to the right.

**d.** A 90% CI is

$$.645 \pm 1.645\sqrt{\frac{(.645)(1-.645)}{124}}$$

which is .645 ± .071, or (.574, .716).

A 99% CI is

$$.645 \pm 2.576\sqrt{\frac{(.645)(1-.645)}{124}}$$

which is .645 ± .111, or (.534, .756).

The higher confidence level produces a wider confidence interval. All of these intervals have the same midpoint: the sample proportion .645.

**e.** Because none of these intervals includes the value .5, it does not appear to be plausible that 50% of all kissing couples lean to the right. In fact, all of the intervals lie entirely above .5, so the data suggest that more than half of all kissing couples lean to the right. The value 2/3 is quite plausible for this population proportion because .667 falls within all three confidence intervals. The value 3/4 is not very plausible because only the 99% CI includes the value .75; the 90% CI and 95% CI do not include .75 as a plausible value.

**f.** The sample size condition is clearly met, as $n\hat{p} = 80$ is greater than 10, and $n(1-\hat{p}) = 44$ is also greater than 10. But the other condition is that the sample be randomly drawn from the population of all kissing couples. In this study, the couples selected for the sample were those who happened to be observed in public places while the researchers were watching. Technically, this is not a random sample, and so you should be cautious about generalizing the results of the confidence intervals to a larger population.