Section 8.4 Confidence Intervals for Proportions (Traditional)

No need to stop at confidence intervals for means. The mechanics are virtually identical for confidence intervals for proportions.

$$p \pm \text{ margin of error } = p \pm z_{\frac{\alpha}{2}} \times \sigma_p = p \pm z_{\frac{\alpha}{2}} \times \sqrt{\frac{p\,(1-p)}{n}}.$$

A confidence interval for the proportion is computed on the TI-83/84 using the **1-PropZInterval** command from statistical tests.

Example 1 *A survey of 128 payday loan borrowers from August finds that 72 have previously used a payday loan. Construct a 99% confidence for the proportion of payday loan borrowers who use the service more than once.*
The point estimate is $p = \frac{72}{128} = 0.562\,5$. *The endpoints of the confidence interval are computed from* $0.5625 \pm 2.575\sqrt{\frac{.5625(1-.5625)}{128}}$. *The lower bound is* $0.5625 - 2.575\sqrt{\frac{.5625(1-.5625)}{128}} = 0.449\,59$. *The upper bound is* $0.5625 + 2.575\sqrt{\frac{.5625(1-.5625)}{128}} = 0.675\,41$. *We are 99% confident that the true proportion of repeat users of payday loans falls in the interval* $(0.449\,59, 0.675\,41)$.

Example 2 *The KSU Dodgeball exploratory committee claims that 35% of KSU students will attend a KSU Dodgeball game at least once during the season. A study group collects data from 100 KSU students where 33 students say they will attend at least one KSU Dodgeball game each season. Has the KSU Dodgeball exploratory committee overestimated the proportion of KSU students who will attend a dodgeball game or is this just chance variation in the sample? Use a 95% confident interval to support your answer.*

The sample percentage is $p = \frac{33}{100} = 0.33$. The endpoints of the confidence interval are computed from $0.33 \pm 1.96\sqrt{\frac{.33(1-.33)}{100}}$. The lower bound is $0.33-1.96\sqrt{\frac{.33(1-.33)}{100}} = 0.237\,84$ an the upper bound is $0.33+1.96\sqrt{\frac{.33(1-.33)}{100}} = 0.422\,16$. We are 95% confident that the true percentage of KSU students who will attend at least one dodgeball game per season is in the interval $(0.237\,84, 0.422\,16)$. We have no evidence to suggest the KSU Dodgeball exploratory committee overestimated since 35% falls in the confidence interval.

Exercise 1 *Suppose we would like our margin of error of* $1.96 * \sqrt{\frac{.33(1-.33)}{100}} = 9.\,216\,2 \times 10^{-2}$ *to be under 5% given the preliminary point estimate of* $p = .33$. *How large of a sample should we collect?*

We wish to solve $1.96 * \sqrt{\frac{.33(1-.33)}{n}} \leq .05$.

$$\begin{aligned}
1.96 * \sqrt{\frac{.33\,(1-.33)}{n}} &\leq .05. \\
\sqrt{\frac{.33\,(1-.33)}{n}} &\leq \frac{.05}{1.96} \\
\frac{.33\,(1-.33)}{n} &\leq \left(\frac{.05}{1.96}\right)^2 \\
.33\,(1-.33) * \left(\frac{1.96}{.05}\right)^2 &\leq n \\
339.75 &\leq n
\end{aligned}$$

In general,

$$n \geq p(1-p)\left(\frac{z_{\frac{\alpha}{2}}}{m}\right)^2$$

where m is the desired bound.

Exercise 2 *Suppose we plan to compute a confidence interval for the proportion of KSU faculty members who own a robot vacuum. Furthermore, we wish to bound the margin of error to within 3% at 95% confidence. How large of a sample should we select? Note that we have no sample estimate p to use. We utilize a value for p that maximizes the sample size. That value is one that maximizes* $p(1-p)$. . *A little calculus shows us that value is* $p = .5$. *Let* $f(p) = p(1-p) = p - p^2$. *We find* $f'(p) = 1 - 2p$. *Setting the derivative equal to 0, we find that the extreme value occurs at* $p = .5$. *Going back to our original problem we find* $n \geq .5 * .5 * \left(\frac{1.96}{.03}\right)^2 = 1067.1$.

Exercise 3 *An experiment observes 177 KSU students as they attempt to solve the same Sudoku puzzle. Only 19 students are able to solve the puzzle. Construct a 99% confidence interval for the true percentage of all KSU students who can solve this particular puzzle.*

Exercise 4 *A random sample of 140 KSU students finds that 37 of those polled live on campus. Construct a 99% confidence interval for the true population proportion of students who live on campus.*

Exercise 5 *Baseball legend Casey (https://en.wikipedia.org/wiki/Casey_at_the_Bat) hit .325 during the regular season. During a playoff game, Casey comes up to bat with the bases loaded and the announcers point out that Casey had 20 hits in 49 at bats with the bases loaded during the regular season. The announcers infer that Casey is better at the plate when the bases are loaded than in general. At a 99% level of confidence, are the announcers correct or is this just an example of chance variation?*

Remark 1 *Our textbook discusses a newer approach that adjusts the number of success from x to $x + 2$ and n to $n + 4$. The TI uses the traditional method. Always know the algorithm and process behind your technology of choice.*

1 Exercises

1. Navidi Section 5.2: 1-5, 13