Confidence Intervals for Population Proportion p: Overview

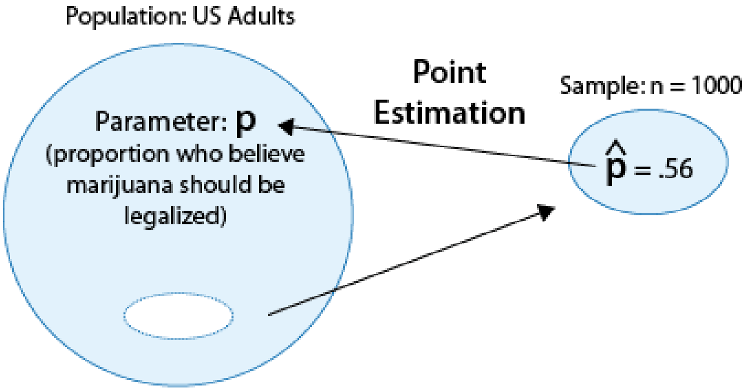
Learning Objective: Explain what a confidence interval represents and determine how changes in sample size and confidence level affect the precision of the confidence interval.

Learning Objective: Find confidence intervals for the population mean and the population proportion (when certain conditions are met), and perform sample size calculations.

Overview

As we mentioned in the introduction to this module, when the variable that we're interested in studying in the population is **categorical**, the parameter we are trying to infer about is the **population proportion (p)** associated with that variable. We also learned that the point estimator for the population proportion p is the sample proportion $\hat{p}$.

To refresh your memory, here is a picture that summarizes an example we looked at.



We are now moving on to interval estimation of p. In other words, we would like to develop a set of intervals that, with different levels of confidence, will capture the value of p. We've actually done all the groundwork and discussed all the big ideas of interval estimation when we talked about interval estimation for μ, so we'll be able to go through it much faster. Let's begin.

Recall that the general form of any confidence interval for an unknown parameter is:

$$estimate \pm margin \;\; of \;\; error$$

Since the unknown parameter here is the population proportion p, the point estimator (as I reminded you above) is the sample proportion $\hat{p}$. The confidence interval for p, therefore, has the form:

$$\hat{p} \pm m$$

(Recall that m is the notation for the margin of error.) The margin of error (m) tells us with a certain confidence what the maximum estimation error is that we are making, or in other words, that $\hat{p}$ is different from p (the parameter it estimates) by no more than m units.

From our previous discussion on confidence intervals, we also know that the margin of error is the product of two components:

$$m = confidence \;\; multiplier \;\; \cdot SD \;\; of \;\; the \;\; estimator$$

To figure out what these two components are, we need to go back to a result we obtained in the Sampling Distributions module of the Probability unit about the sampling distribution of $\hat{p}$. We found that under certain conditions (which we'll come back to later), $\hat{p}$ has a normal distribution with mean p, and standard deviation $\sqrt{\frac{p(1-p)}{n}}$. This result makes things very simple for us, because it reveals what the two components are that the margin of error is made of:

- Since, like the sampling distribution of $\overline{X}$, the sampling distribution of $\hat{p}$ is normal, the confidence multipliers that we'll use in the confidence interval for p will be the same z* multipliers we use for the confidence interval for μ when σ is known (using **exactly** the same reasoning and the same probability results). The multipliers we'll use, then, are: **1.645, 2, and 2.576 at the 90%, 95% and 99% confidence levels, respectively.**

- The standard deviation of our estimator $\hat{p}$ is $\sqrt{\frac{p(1-p)}{n}}$

Putting it all together, we find that the confidence interval for p should be: $\hat{p} \pm z^* \cdot \sqrt{\frac{p(1-p)}{n}}$. We just have to solve one practical problem and we're done. We're trying to estimate the **unknown** population proportion **p**, so having it appear in the confidence interval doesn't make any sense. To overcome this problem, we'll do the obvious thing ...

We'll replace p with its sample counterpart, $\hat{p}$, and work with the **standard error of** $\hat{p}$, $\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$.

Now we're done. The confidence interval for the population proportion p is:

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

As you'll see from the examples we'll present in this unit, estimating the population proportion comes up a lot in the context of polls.