1

CONFIDENCE INTERVALS POPULATION PROPORTIONS

Unit 4A - Statistical Inference Part 1

UF UNIVERSITY of FLORIDA

{Confidence Intervals for Population Proportions}

Now that we have discussed the theory of confidence intervals in the case of population means, we can quickly go through the same logic for population proportions.

2

Confidence Interval for p

- Assumptions:
  - Simple Random Sample from the Population
  - Binomial Properties are satisfied for the sample

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

$$n\hat{p} \geq 10 \text{ and } n(1-\hat{p}) \geq 10$$

UF UNIVERSITY of FLORIDA

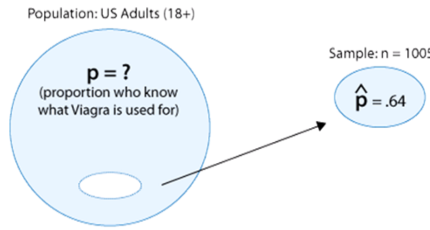
You can review the materials for more detailed discussion of the development of this equation.

In order to apply the results we learned about the sampling distribution of p-hat, we must have a simple random sample from the population and the binomial properties must be satisfied – we have a fixed number of trials, the true probability of success, p, is the same for each trial, and the trials are independent of each other.

You may notice that in the formulas presented here, we have substituted p-hat for p in the equation for the standard error and for the sample size requirement. Since we don't know the true value of p, we must estimate it from our data, using p-hat, in order to perform these calculations.

3

Example



$$\hat{p} \pm 1.96 \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.64 \pm 1.96 \cdot \sqrt{\frac{0.64(1-0.64)}{1005}}$$

$$= 0.64 \pm 0.03$$

$$= (0.61, 0.67)$$

UF UNIVERSITY of FLORIDA

Here is an example where a sample of 1105 adults were asked if they were aware that Viagra was an impotency medication. 643 or approximately 64% of respondents stated they were aware.

Substituting for n, p-hat, and z* in our equation we calculate the margin of error to be 3% and the final confidence interval to be from 61% to 67%.

Thus, we can be 95% confident that the proportion of all U.S. adults (18+) who were already familiar with Viagra by that time was between 0.61 and 0.67 (or 61% and 67%)

The fact that the margin of error equals 0.03 says we can be 95% confident that the unknown population proportion *p* is within 0.03 (3%) of the observed sample proportion 0.64 (64%). In other words, we are 95% confident that 64% is "off" by no more than 3%.

Sample Size Needed

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

$$n = \frac{(z^*)^2}{4 \cdot m^2}$$

UF UNIVERSITY of FLORIDA

The required sample size needed to estimate an unknown population proportion with a specified margin of error is found in a similar manner and is given by this equation.

As we usually don’t have an estimate for the true proportion, we can use ½ as the “worse case scenario” and simplify our equation to the second one in that case.

ER Example

Obese				
obese	Frequency	Percent	95% Confidence Limits for Percent	
No	269	61.5561	56.9772	66.1350
Yes	168	38.4439	33.8650	43.0228
Total	437	100.000		

UF UNIVERSITY of FLORIDA

Here is an example from software.

The data are on a random sample of 437 emergency room patients.

Based upon our sample, with 95% confidence we estimate that between 34% and 43% of ER patients are obese.

Suppose we wish to determine if the true prevalence of obesity in the population of ER patients is different from that in the state overall, which is 27%. What can we say?

Since 27% is far below the range of our confidence interval, based upon our data, 27% is not a plausible value for the true prevalence of obesity in the population of ER patients and thus the prevalence of obesity in the population is significantly different from that in the state overall.

CONFIDENCE INTERVALS POPULATION PROPORTIONS

UF UNIVERSITY of FLORIDA

That concludes our discussion of confidence intervals for now. We will continue to use confidence intervals to estimate unknown parameters in future analyses, however, we will rely on software to calculate them for us and focus on interpreting the results in context.

For problems like those discussed here it is important to first choose whether the variable under study is categorical or quantitative, then to determine whether the methods can be used.

When interpreting confidence intervals, it is extremely important to **mention the confidence level** (**avoid using** the term **probability** or **chance** in the interpretation) and specify exactly **what parameter** is being estimated about **which population**.