1342-Notes_Navidi_8-3_confidence-intervals-proportion

Monday, November 18, 2019 9:40 AM

1342-Notes_Navidi_8-3_confidence-intervals-proportion

8.3: Confidence Intervals for One Population Proportion

Point estimates for the population proportion:

Recall:

Definition: A *point estimate* is the value of a statistic that estimates the value of a parameter.

Definition: A confidence interval for an unknown parameter is an interval of numbers generated by a point estimate for that parameter.

Definition: The *confidence level* (usually given as a percentage) represents how confident we are that the confidence interval contains the parameter.

If a large number of samples is obtained, and a separate point estimate and confidence interval are generated from each sample, then a 95% confidence level indicates that 95% of all these confidence intervals contain the population parameter.

A confidence interval is obtained by placing a *margin of error* on either side of the point estimate of the parameter.

In other words, the confidence interval consists of: Point estimate $\pm$ margin of error

The point estimate of the population proportion p is the sample proportion $\hat{p}$.
The point estimate of the mean of the sampling distribution of the sample proportions is $\mu_{\hat{p}} = \hat{p}$.
The point estimate of the standard deviation of the sampling distribution of the sample proportions is

$$\sigma_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}.$$

estimate of the standard error

So, for every sample, the sample proportion will be in the center of the confidence interval. If we use E to indicate the margin of error, the confidence interval is $\hat{p} \pm E$, or $(\hat{p}-E, \hat{p}+E)$

If we use the sample proportion $\hat{p}$ as a starting point, we should be able to write the confidence interval as $(\hat{p}-z_c\sigma_{\hat{p}}, \hat{p}+z_c\sigma_{\hat{p}})$, where $\sigma_{\hat{p}}$ is the standard deviation of the sampling distribution of the sample proportions, and z_c is a multiplier that tells us how many standard deviations (of the sampling distribution of the sample proportions) lie between the sample proportion $\hat{p}$ and the edge of the confidence interval. We call this z_c the *critical value* for a z-score in the sampling distribution of the sample proportions.

Constructing the confidence interval for the proportion:

Procedure:

1. Verify that $np \geq 10$ and $n(1-p) \geq 10$ and that the sample is no more than 5% of the population. (if we don't have p, use $n\hat{p} \geq 10$, $n(1-\hat{p}) \geq 10$)
2. Determine the confidence level, $1-\alpha$.
3. Determine the critical value $z_{\alpha/2}$ (using the standard normal table).
   $z_{\alpha/2}$
4. Use the sample proportion to estimate the standard deviation of the sampling distribution of the sample proportions:

   $$\sigma_{\hat{p}} \approx \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}.$$ (estimate the standard error)

5. Multiply the critical value $z_{\alpha/2}$ by the estimated standard deviation $\sigma_{\hat{p}} \approx \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ to obtain the margin of error.
6. Add and subtract the margin of error from the sample proportion to obtain the lower and upper bounds of the confidence interval:

   Lower bound: $\hat{p} - z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

   Upper bound: $\hat{p} + z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

Useful critical values of z:
(You can use these instead of looking in the table every time).

Confidence Level	α	Area in each tail, $\alpha/2$	Critical value $z_{\alpha/2}$
90%	0.10	0.05	1.645
95%	0.05	0.025	1.96
99%	0.01	0.005	2.575

Example 1: In a random sample of 537 Americans, 173 indicated that they frequently ate peanut butter. Construct and interpret the 90% and the 95% confidence intervals for the proportion of Americans who frequently eat peanut butter.

Check assumptions:
Is sample big enough for $\hat{p}$ to be normally distributed?

$n = 537$

eat PB often	don't eat PB often
173	364

$n - x = 537 - 173$

Both ≥ 10, so $\hat{p}$ normally distributed.

Sample info

$n = 537$

$x = 173$ = # in sample who eat PB often

$\hat{p} = \frac{x}{n} = \frac{173}{537} \approx 0.322$

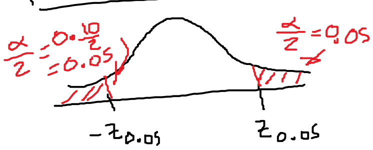
$1 - \hat{p} = 1 - 0.322 = 0.678$

Find std error:

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

$$= \sqrt{\frac{0.322(0.678)}{537}} \approx 0.02016$$

For 90% CI: critical value: $z_{0.05} = 1.645$ from z-table



see next page

Sample size needed to estimate the population proportion within a given margin of error:

When constructing a confidence interval about the sample proportion $\hat{p}$, the margin of error is

$$z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}.$$

We can solve this for n:

In order to calculate the n needed, we need an educated guess for the population proportion p. If such an educated guess is available (perhaps from a prior study), we can use the above formula to calculate n. If not, we use the very conservative assumption that $\hat{p} = 0.5$, which gives us the maximum possible value for $\hat{p}(1-\hat{p})$, which is (0.5)(0.5) = 0.25.

<u>Required sample size for estimation of the population proportion</u>:

a) For a specified α associated with a confidence level, the sample size required to estimate the population proportion within a margin of error E is

$$n = \hat{p}_g(1-\hat{p}_g)\left(\frac{z_{\alpha/2}}{E}\right)^2,$$

where $\hat{p}_g$ is an educated guess for the population proportion p.

b) If you know a likely range of values for the sample proportion, choose the value in that range that is closest to 0.5. Use this value as the educated guess $\hat{p}_g$ in the above formula. (The above formula will be at its maximum when $\hat{p}_g = 0.5$. Thus a larger sample is required when $\hat{p}_g$ is close to 0.5, compared to when $\hat{p}_g$ is further away. To be sure we have a big enough sample, we look at all the possible values for $\hat{p}$ and choose the one closest to 0.5.)

c) If no estimate for the population proportion is available, we should use a sample size of at least

$$n = 0.25\left(\frac{z_{\alpha/2}}{E}\right)^2.$$

In all cases, because the calculated n is considered a minimum threshold, we round the calculated value of n <u>up</u> to the nearest whole number *above*.

<u>Example 2:</u> A pollster wishes to estimate the percentage of likely voters who support Candidate A. Based on earlier polls, the pollster expects the candidate's level of support to be approximately 38%. What sample size should be obtained if the pollster wishes to estimate the candidate's support level within a margin of error of 3 percentage points, with 95% confidence?

Example 3: a) Estimate the minimum sample size required to estimate the population proportion within a margin of error of 0.02, if the proportion is expected to be between 0.1 and 0.3. Use a confidence level of 95%.
b) Suppose the sample size from part (a) is obtained, and that the proportion of the characteristic of interest turns out to be 0.25. Construct the 95% confidence interval for the population proportion. What is the margin of error?

Example 4: Estimate the minimum sample size required to estimate the population proportion within a margin of error of 0.03, if you have no idea what the proportion will turn out to be.

Ex 1 cont'd:

For 90% CI

Confidence interval: $\hat{p} \pm z_{0.05} \cdot \sigma_{\hat{p}}$

use stored value from calculator

Upper Bound: $0.322 + 1.645(0.02163)$

$= 0.355$

Lower Bound: $0.322 - 1.645(0.02163)$

$= 0.289$

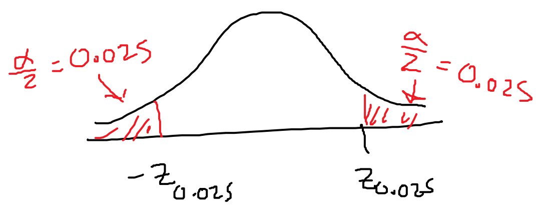
$\boxed{90\%\ \text{CI}: (0.289, 0.355)}$

For 95% CI:

Find critical value

$\alpha = 0.05$

$z_{0.025} = 1.96$



$=$ Look up Area = 0.975 in table $\Rightarrow z_{0.025} = 1.96$

$\hat{p} \pm z_{0.025} \cdot \sigma_{\hat{p}}$

Upper Bound: $0.322 + 1.96(0.02163)$

$= 0.362$

Lower Bound: $0.322 - 1.96(0.02163)$

$= 0.282$

$\boxed{95\%\ \text{CI}: (0.282, 0.362)}$