Chapter 9-10

Confidence Intervals and Hypothesis Testing

CI for Single Proportion

Statistics

Mehmet Güray Güler, PhD

Last updated 21.07.2020

The Confidence Interval

- Using CLT $Z = \frac{\hat{P}-p}{\sqrt{p(1-p)/n}}$ is approximately normally distributed.
- Hence an approximate CI is $P\left(\hat{P} - z_{a/2}\sqrt{\frac{p(1-p)}{n}} < p < \hat{P} + z_{a/2}\sqrt{\frac{p(1-p)}{n}}\right) = 1 - a.$
- Compare with
- $P\left(\bar{X} - z_{\alpha/2}\,\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\alpha/2}\,\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$
- We don't know the value of *p*, the population proportion
  - Use its estimate $\hat{p}$ to find an **approximate** CI

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

Example – TiViBu Subscription

- A random sample of 500 households was taken from among all the TV owners in İstanbul.
- The sample contained 340 subscribers to TiViBu service.
- Find a 95% CI for the true proportion p of households with TV sets in İstanbul that have subscription to TiViBu services.
- The variable in this problem is binary, X, defined as:

$$X = \begin{cases} 1, & \text{if the household is a subscriber,} \\ 0, & \text{otherwise} \end{cases}$$

Example – TiViBu Subscription

- DATA: $X_1, X_2, \ldots, X_{500}$. We are given that $\Sigma X_\text{i} = 340$.
- p = the proportion of Tivibu subscribers in İstanbul.
- $\hat{p} = 340/500 = 0.68$.

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

- A 95% CI for p:

$$= 0.68 \pm 1.96\sqrt{\frac{(0.68)(0.32)}{500}}$$

$$=> 0.64 < p < 0.72.$$

Error in Estimating p by $\hat{\boldsymbol{p}}$

- The error: the (absolute) difference between the point estimate and the (unknown) true value is called the error in estimating the parameter.

$$e = z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}$$

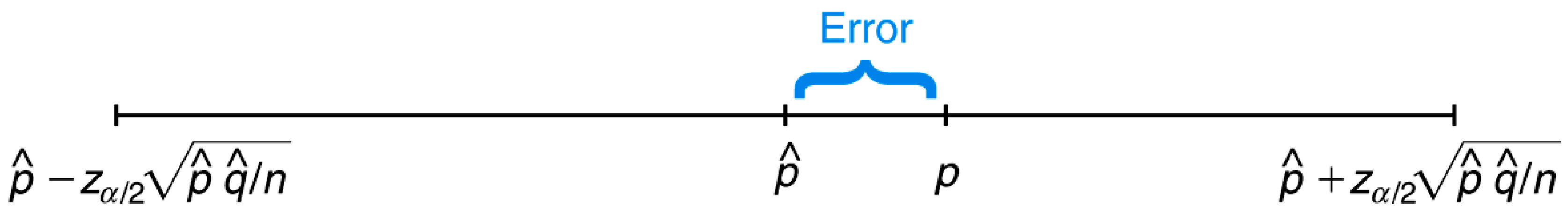


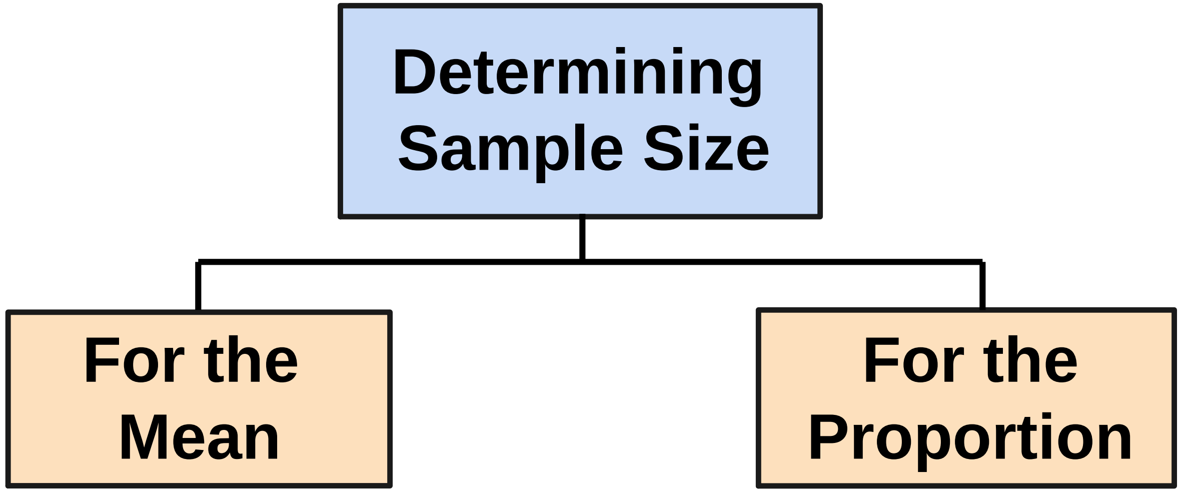
Figure 1. Confidence interval and the size of the error in estimating p.

Error in Estimating p by $\hat{\boldsymbol{p}}$

- Theorem. We are (1−α) 100% confident that the error *e* in estimating the population proportion **p** by the sample proportion can be at most equal to $e = z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}$

- Q: What is the size of the error in Example 4?

The Needed Sample Size



$$n = \left(\frac{z_{a/2}\, s}{e} \right)^2 .$$

$$n = \frac{z_{a/2}^2\, \hat{p}(1 - \hat{p})}{e^2} .$$

e = margin of error = sampling error

The Needed Sample Size

- THEOREM. We can be $(1-\alpha)$ 100% confident that the error in estimating the population proportion *p* by using the sample proportion will not exceed a specified amount *e* when the sample size is taken as: $n = \frac{z_{\alpha/2}^2 \hat{p}(1-\hat{p})}{e^2}$
- The value of $\hat{p}$? Two approaches:
  1. Take its value to make *n* maximum, this means $\hat{p} = 0.5$, or
  2. Take a preliminary sample of size $n \geq 30$ to get an estimate of *p* from the sample, and find *n* needed.

Example – The Needed Sample Size

- *How large a sample is needed if we want to be 95% confident that our estimate of p in Example 4 (Tivibu subscription) has an error within 0.02 of the true value?*
- **Second Approach:** Let’s assume that the first sample of 500 households is a preliminary sample to estimate the needed sample size *n*.
- We then find an initial estimate of *p* as $\hat{p} = 0.68$ in Ex 4.

Example – The Needed Sample Size

- The sample size n to estimate p with an absolute error of 0.02:

$$n = \frac{z_{a/2}^2 \hat{p}(1 - \hat{p})}{e^2} = \frac{(1.96)^2(0.68)(0.32)}{(0.02)^2} » 2090.$$

- Hence we need to take a second sample of 1590 households.

Example – The Needed Sample Size

- first approach. That means we use $p=0.50$ and estimate n as:

$$n = \frac{(1.96)^2(0.5)(0.5)}{(0.02)^2} » 2401.$$

- Hence taking a preliminary sample may reduce the cost by avoiding too large a sample for the desired precision.