STAT 108 Lecture 15 Tu, Oct 27, 2015

8.2 Estimation of a Population Mean μ When σ is Assumed Known

Recall that a critical value $z_{\alpha/2}$ for a $100\cdot(1-\alpha)\%$ CI is defined as satisfying $P(-z_{\alpha/2} \le z \le z_{\alpha/2}) = 1-\alpha$ where z is a standard normal random variable.

To derive the formula for a $100\cdot(1-\alpha)\%$ CI for a population mean μ assuming that population standard deviation σ is known, we consider the case of a large sample size ($n \geq 30$). By the CLT, the distribution of $\overline{x}$ is approximately normal with mean μ and standard deviation $\sigma/\sqrt{n}$. We can write

$$P\left(-z_{\alpha/2} \le \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \le z_{\alpha/2}\right) = 1-\alpha,$$

or, equivalently,

$$P\left(-z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \bar{x}-\mu \le z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha,$$

or, finally,

$$P\left(\bar{x}-z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x}+z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

Definition. A $100 \cdot (1-\alpha)\%$ CI for μ when σ is known is

$$\left[\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\ \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

Example. Suppose it is known that the standard deviation of the commute time to school is 11 minutes. A random sample of 100 students produces the sample mean commute time of 40 minutes. Compute the 95% CI for the population mean commute time.

Solution.

$$\bar{x} \pm z_{0.025}\frac{\sigma}{\sqrt{n}} = 40 \pm 1.96\frac{11}{\sqrt{100}}$$

$$= 40 \pm 2.2 = [37.8,\ 42.2]$$

Example. Suppose it is know that the standard deviation of the amount of money college students spend on textbooks is \$57. A sample of 130 students resulted in an average of \$422 spent on textbooks. Give a 90% confidence interval for the mean amount of money spent by college students on textbooks.

Solution.

$$\bar{x} \pm z_{0.05}\frac{\sigma}{\sqrt{n}} = 422 \pm 1.645\frac{57}{\sqrt{130}}$$

$$= 422 \pm 8.22 = [413.78,\ 430.22]$$

How to determine a required sample size for a desired margin of error?

Note that for a $100\cdot(1-\alpha)\%$ CI for μ the margin of error is equal to

$$m = z_{\alpha/2}\frac{\sigma}{\sqrt{n}}.$$

Suppose we are given α, σ and *m*, and we need to find *n*.

We have to solve with respect to *n* the equation $m = z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}.$

We have $\sqrt{n} = z_{\alpha/2}\dfrac{\sigma}{m}$, or $n = \left(\dfrac{z_{\alpha/2}\sigma}{m}\right)^2.$

In practice, *n* is the smallest integer such that

$$n \geq \left(\frac{z_{\alpha/2}\sigma}{m}\right)^2.$$

Note that the margin of error in our textbook is denoted by E and the formula for n is

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

Example. Suppose it is known that the standard deviation of the commute time to school is 11 minutes. Suppose also that we want to compute a 95% CI for μ with the margin of error not exceeding 3 minutes. What should the sample size be?

<u>Solution</u>. The required sample size n is the smallest integer such that

$$n \geq \left(\frac{z_{\alpha/2}\sigma}{m} \right)^2 = \left(\frac{(1.96)(11)}{3} \right)^2 = 51.65,$$

that is, $n = 52$.

Note that for n=52, $m = (1.96)(11)/\sqrt{52} = 2.99 < 3,$

whereas for n=51, $m = (1.96)(11)/\sqrt{51} = 3.02 > 3.$

<u>Example</u>. Suppose it is know that the standard deviation of the amount of money college students spend on textbooks is \$57. Suppose we want to construct a 90% confidence interval for the population mean, and we want the maximum error of estimate to be within \$15 of the population mean. Find the minimum required sample size.

Solution. The required sample size n is the smallest integer such that

$$n \geq \left(\frac{z_{\alpha/2}\sigma}{m} \right)^2 = \left(\frac{(1.645)(57)}{15} \right)^2 = 39.075,$$

that is, $n = 40$.

Note that for n=40, $m = (1.645)(57)/\sqrt{40} = 14.83 < 15,$

whereas for n=39, $m = (1.645)(57)/\sqrt{39} = 15.01 > 15.$

8.4 Estimation of a Population Proportion: Large Samples

Recall that when n is large ($n \geq 30$), then the sample proportion $\hat{p}$ is approximately normally distributed with mean p and standard deviation $\sqrt{\frac{p(1-p)}{n}}$.

Definition. A $100 \cdot (1-\alpha)\%$ CI for p is

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

Note that in this formula we estimated the unknown p in the standard deviation by $\hat{p}$.

Example. In a random sample of 50 homeowners, 38% don't like their neighbors. Find a 99% CI for the true population proportion of homeowners who don't like their neighbors.

Solution. It is given that

$$n = 50 \text{ and } \hat{p} = 0.38.$$

Therefore, a 99% CI for *p* is

$$0.38 \pm 2.575 \sqrt{\frac{(0.38)(1-0.38)}{50}}$$
$$= 0.38 \pm 0.177 = [0.203,\ 0.557].$$

Example. Of 200 high school students surveyed randomly, 179 said that they get weekly allowance from their parents. Construct a 95% CI for p.

Solution. It is known that

$$n = 200, \text{ and } \hat{p} = \frac{179}{200} = 0.895.$$

Thus, a 95% CI for p is

$$0.895 \pm 1.96\sqrt{\frac{(0.895)(1-0.895)}{200}}$$
$$= 0.895 \pm 0.042 = [0.853, 0.937].$$

Example. A CEO wants to know what proportion of the company's employees are routinely late for work. A sample of 70 employees revealed that 14% are routinely late for work. Compute a 90% confidence interval for the true proportion of employees who are routinely late for work.

Solution. The 90% CI for p is

$$0.14 \pm 1.645 \sqrt{\frac{(0.14)(1-0.14)}{70}}$$

$$= 0.14 \pm 0.064 = [0.076, 0.204].$$

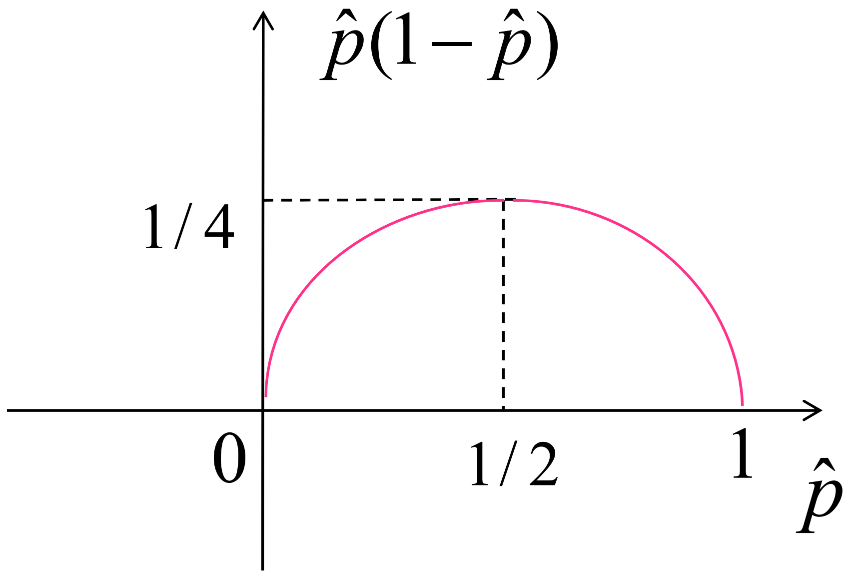
How to determine a required sample size for a desired margin of error?

Suppose we are given the confidence level α and the margin of error m of a CI for p. We are asked to compute the required sample size n.

For a $100 \cdot (1-\alpha)\%$ CI for *p*, the margin of error is

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

Here we know m and $z_{\alpha/2}$, and need to find *n*. But we don't know the value of $\hat{p}$, because we haven't drawn the sample yet. This is what we do:



From the picture, $\hat{p}(1-\hat{p}) \leq \frac{1}{4}$.

Thus,

$$z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \le z_{\alpha/2}\sqrt{\frac{1/4}{n}} = \frac{z_{\alpha/2}}{2\sqrt{n}}.$$

The margin of error $m = \frac{z_{\alpha/2}}{2\sqrt{n}}$ is the largest it can be, and the estimator of n is called the **most conservative**. The value of n is taken to be the smallest integer such that

$$n \ge \left(\frac{z_{\alpha/2}}{2m}\right)^2.$$

<u>Example</u>. Find the most conservative estimate of the sample size required to estimate, with a 99% level of confidence, the true population proportion of homeowners who don't like their neighbors to within 5%.

Solution. The most conservative estimate of n is the smallest integer such that

$$n \geq \left(\frac{2.575}{(2)(0.05)} \right)^2 = 663.06$$

or $n = 664$.

Example. A CEO wants to know what proportion of the company's employees are routinely late for work. What is the most conservative estimate of the sample size that would limit the margin of error to within ± 0.1 of the population proportion for a 90% confidence interval?

Solution. Given $m = 0.1$, and $z_{0.05} = 1.645$. Hence, the most conservative estimate of n is the smallest integer such that

$$n \geq \left(\frac{1.645}{(2)(0.1)} \right)^2 = 67.65$$

or $n = 68$.