7.1: Estimation of a Population Proportion

Learning Objectives

- To understand how to apply the formula for a confidence interval for a population proportion.

Since from Section 6.3, we know the mean, standard deviation, and sampling distribution of the sample proportion $\hat{p}$, the ideas of the previous two sections can be applied to produce a confidence interval for a population proportion. Here is the formula.

Large Sample $100(1-\alpha)\%$ Confidence Interval for a Population Proportion

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

A sample is large if the interval $[p-3\sigma_{\hat{p}}, p+3\sigma_{\hat{p}}]$ lies wholly within the interval $[0,1]$.

In actual practice the value of p is not known, hence neither is $\sigma_{\hat{p}}$. In that case we substitute the known quantity $\hat{p}$ for p in making the check; this means checking that the interval

$$\left[\hat{p}-3\sqrt{\frac{\hat{p}(1-\hat{p})}{n}},\ \hat{p}+3\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]$$

lies wholly within the interval $[0,1]$.

Example 7.1.1

To estimate the proportion of students at a large college who are female, a random sample of 120 students is selected. There are 69 female students in the sample. Construct a 90% confidence interval for the proportion of all students at the college who are female.

Solution

The proportion of students in the sample who are female is

$$\hat{p} = 69/120 = 0.575$$

Confidence level 90% means that $\alpha = 1-0.90 = 0.10$ so $\alpha/2 = 0.05$. From the last line of Figure 7.1.6 we obtain $z_{0.05} = 1.645$.

Thus

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.575 \pm 1.645\sqrt{\frac{(0.575)(0.425)}{120}} = 0.575 \pm 0.074$$

One may be 90% confident that the true proportion of all students at the college who are female is contained in the interval $(0.575-0.074, 0.575+0.074) = (0.501, 0.649)$.

Summary

- We have a single formula for a confidence interval for a population proportion, which is valid when the sample is large.
- The condition that a sample be large is not that its size n be at least 30, but that the density function fit inside the interval $[0,1]$.