§7.4 Large-Sample Confidence Intervals for a Population Proportion.

Population proportion is related to binomial.

Example 40 statistic students are surveyed and 23 of them said the last exam was difficult. What fraction of all statistics students has this same opinion?

Restated as a binomial, if we randomly selected one of the 40 students and asked if they thought the test was hard, then $\hat{p} = \frac{23}{40}$ and $\hat{q} = \frac{17}{40}$. Here we use $\hat{p}$ (read as "p hat") to denote the probability of a success from the sample. We are then asked to infer what p for the population would be.

Definition (1) $\hat{p}$ from the large sample is the point estimator for p, the population proportion. $\hat{p}$ is called the unbiased estimator of p.

(2) the standard deviation of the sample distribution of $\hat{p}$ is $\sigma_{\hat{p}} = \sqrt{pq/n}$ where $q = 1-p$.

(3) the sample distribution of $\hat{p}$ is approximately normal, provided the sample size is large. that is $n\hat{p} \geq 15$ and $n\hat{q} \geq 15$.

Definition for large-sample confidence intervals for p, the margin of error is $z_{\alpha/2}\sigma_{\hat{p}} = (z_{\alpha/2})\left(\sqrt{\frac{pq}{n}}\right) \approx (z_{\alpha/2})\left(\sqrt{\frac{\hat{p}\hat{q}}{n}}\right)$. That is the confidence interval for a population proportion is

$$\hat{p} \pm (z_{\alpha/2})\left(\sqrt{\frac{\hat{p}\hat{q}}{n}}\right)$$

Where $\hat{p} = \frac{x}{n}$ and $\hat{q} = 1 - \hat{p}$.

In the above example $n = 40$, $x = 23$, $\hat{p} = \frac{23}{40}$, $\hat{q} = 1 - \frac{23}{40} = \frac{17}{40}$. Note $n\hat{p} = 23 \geq 15$ and $n\hat{q} = 17 \geq 15$. So the 95% confidence interval for p would be:
$\alpha = .05$, $z_{.025} = 1.960$, so

$$\hat{p} \pm (z_{.025})\left(\sqrt{\frac{\hat{p}\hat{q}}{n}}\right) = \frac{23}{40} \pm (1.960)\left(\sqrt{\frac{\left(\frac{23}{40}\right)\left(\frac{17}{40}\right)}{40}}\right)$$

$\Rightarrow$ $(0.422, .728)$ or $0.422 < p < 0.728$.

Note: $\hat{p} = \frac{23}{40} \approx .575$

Note: Since .5 is in the confidence interval we could not conclude that more than half of the statistic students thougt the exam was hard.

If $n\hat{p} < 15$ or $n\hat{q} < 15$, then we can not depend on this proceedure to generate a confidence interval for p.