Introduction to Statistics Note

2024 Spring Semester

21 CST H3Art

Chapter 8: Inference for Proportions

8.1 Inference for a Single Proportion

We can construct a confidence interval for an unknown population proportion p:

$$\textbf{statistic} \pm (\textbf{critical value}) \cdot (\textbf{standard deviation of statistic})$$

The sample proportion $\hat{p}$ is the statistic we use to estimate p. When the independent condition is met, the **standard deviation** of the sampling distibution of $\hat{p}$ is:

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

Since we don't know p, we replace it with the sample proportion $\hat{p}$. This gives us the **standard error (SE)** of the sample proportion:

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

Once we find the critical value z^*, our **confidence interval for the population proportion** p is:

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

The z statistic has approximately the standard Normal distribution when H_0 is true. P-values therefore come from the standard Normal distribution. Here is a summary of the details for a z **test for a proportion**, to test the hypothesis $H_0 : p = p_0$, compute the z statistic:

$$z = \frac{\hat{p} - p}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

To do Normal calculations, the sample size value should **at least** 10.

margin of error:

$$m = z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

To **determine the sample size** n that will yield a level C confidence interval for a population proportion p with a maximum margin of error, solve the following:

$$n = \left(\frac{z^*}{m}\right)^2 p^*(1-p^*)$$

where p^* is a **guessed value for the sample proportion**. The margin of error will always be less than or equal to m if you take the guess p^* to be 0.5.

8.2 Comparing Two Proportions

Sampling distribution of $\hat{p_1} - \hat{p_2}$, its standard deviation is:

$$\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

Because we don't know the values of the parameters p_1 and p_2, we replace them in the standard deviation formula with the sample proportions. The result is the **standard error** of the statistic $\hat{p}_1 - \hat{p}_2$:

$$\mathbf{SE} = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

When the **Random** and **Normal** conditions are met, an approximate level C confidence interval for $(\hat{p}_1 - \hat{p}_2)$ is:

$$(\hat{p}_1 - \hat{p}_2) \pm z^* \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

- **Random**: The data are produced by a **random sample** of size n_1 from Population 1 and a **random sample** of size n_2 from Population 2 or by two groups of sizes n_1 and n_2 in a randomized experiment.
- **Normal**: The counts of "successes" and "failures" in each sample or group——$n_1\hat{p}_1$, $n_1(1-\hat{p}_1)$, $n_2\hat{p}_2$ and $n_2(1-\hat{p}_2)$ are **all at least** 10.

Significance Test（显著性测试） for Comparing Proportions: To do a test, standardize $\hat{p_1} - \hat{p_2}$ to get a z statistic:

$$\textbf{test statistic} = \frac{\textbf{statistic} - \textbf{parameter}}{\textbf{standard deviation of statistic}}$$
$$z = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{\textbf{standard deviation of statistic}}$$

This **pooled (or combined) sample proportion** is:

$$\hat{p} = \frac{\textbf{count of successes in both samples combined}}{\textbf{count of individuals in both samples combined}}$$

To test the hypothesis $H_0 : p_1 - p_2 = 0$, first find the pooled proportion $\hat{p}$ of successes in both samples combined. Then compute the z statistic:

$$z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Note that the **standard deviation of pooled sample** is:

$$\sigma_{\hat{p}} = \sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$