Difference in Proportions

Grinnell College

October 31, 2025

Warm-up

- Critical value is to confidence as t-statistic is to p-value. Elaborate
- Explain how p-values can be used to perform hypothesis testing
- What are the two components of a t-statistic that can be used to quantify "strength of evidence"?

Differences in Proportion

A useful extension to the distribution of a sample proportion is the *difference in proportions* which, according to the CLT, also follows an approximately normal distribution:

$$\hat{p}_1 - \hat{p}_2 \sim N\left(p_1 - p_2, \ \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}\right)$$

The t-statistic we use implicitly assumes that the true difference is equal to zero:

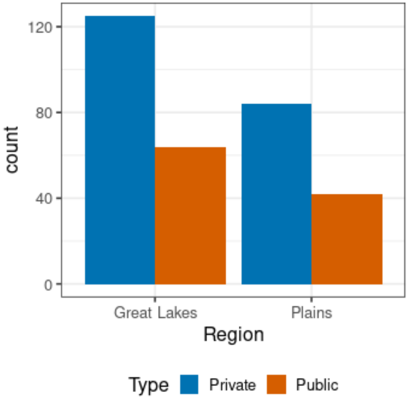
$$t = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}}$$

which follows a t-distribution with $df = n_1 + n_2 - 2$ degrees of freedom

Difference in Proportions

Suppose we are interested in determining if the composition of public and private schools is the same between the Plains region and the Great Lakes

	Private	Public
Great Lakes	125	64
Plains	84	42



Difference in Proportions

	Private	Public	Sum
Great Lakes	125	64	189
Plains	84	42	126

- $\hat{p}_1 = 0.661$
- $\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} = 0.0011$
- $\hat{p}_2 = 0.666$
- $\frac{\hat{p}_2(1-\hat{p}_2)}{n_2} = 0.0017$

Using $C = 1.649$ as our critical value for $df = 313$, we find a 90% CI of

$$\hat{p} \pm C \times \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} = (-0.094, 0.084)$$

Computing t-statistics

	Private	Public	Sum
Great Lakes	125	64	189
Plains	84	42	126

- $\hat{p}_1 = 0.661$
- $\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} = 0.0011$
- $\hat{p}_2 = 0.666$
- $\frac{\hat{p}_2(1-\hat{p}_2)}{n_2} = 0.0017$

Together, this gives us a t-statistic of

$$t = \frac{0.661 - 0.666}{\sqrt{0.0011 + 0.0017}} = -0.09$$

Indicating that our observed data is very near what we would expect if these proportions were truly equal

Key Takeaways

- Proportions share the same properties as the mean
- CLT for proportion and difference of proportion
- Confidence intervals and test statistics computed the same way