5.5 - Hypothesis Testing for Two-Sample Proportions

We are now going to develop the hypothesis test for the difference of two proportions for independent samples. The hypothesis test follows the same steps as one group.

These notes are going to go into a little bit of math and formulas to help demonstrate the logic behind hypothesis testing for two groups. If this starts to get a little confusion, just skim over it for a general understanding! Remember we can rely on the software to do the calculations for us, but it is good to have a basic understanding of the logic!

We will use the sampling distribution of $\hat{p}_1 - \hat{p}_2$ as we did for the confidence interval.

For a test for two proportions, we are interested in the difference between two groups. If the difference is zero, then they are not different (i.e., they are equal). Therefore, the null hypothesis will always be:

$$H_0 \colon p_1 - p_2 = 0$$

Another way to look at it is $H_0 \colon p_1 = p_2$. This is worth stopping to think about. Remember, in hypothesis testing, we assume the null hypothesis is true. In this case, it means that p_1 and p_2 are equal. Under this assumption, then $\hat{p}_1$ and $\hat{p}_2$ are both estimating the same proportion. Think of this proportion as p^*.

Therefore, the sampling distribution of both proportions, $\hat{p}_1$ and $\hat{p}_2$, will, under certain conditions, be approximately normal centered around p^*, with standard error $\sqrt{\dfrac{p^*(1-p^*)}{n_i}}$, for $i = 1, 2$.

We take this into account by finding an estimate for this p^* using the two-sample proportions. We can calculate an estimate of p^* using the following formula:

$$\hat{p}^* = \frac{x_1 + x_2}{n_1 + n_2}$$

This value is the total number in the desired categories $(x_1 + x_2)$ from both samples over the total number of sampling units in the combined sample $(n_1 + n_2)$.

Putting everything together, if we assume $p_1 = p_2$, then the sampling distribution of $\hat{p}_1 - \hat{p}_2$ will be approximately normal with mean 0 and standard error of $\sqrt{p^*(1-p^*)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$, under certain conditions.

Therefore,

$$z^* = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{\sqrt{\hat{p}^*(1-\hat{p}^*)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

...will follow a standard normal distribution.

Finally, we can develop our hypothesis test for $p_1 - p_2$.

Hypothesis Testing for Two-Sample Proportions

Null:

$H_0 \colon p_1 - p_2 = 0$

Conditions:

$n_1\hat{p}_1$, $n_1(1-\hat{p}_1)$, $n_2\hat{p}_2$, and $n_2(1-\hat{p}_2)$ are all greater than five

Test Statistic:

$$z^* = \frac{\hat{p}_1 - \hat{p}_2 - 0}{\sqrt{\hat{p}^*(1-\hat{p}^*)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

> ...where $\hat{p}^* = \dfrac{x_1 + x_2}{n_1 + n_2}$.
>
> The critical values, p-values, and decisions will all follow the same steps as those from a hypothesis test for a one-sample proportion.

‹ 5.4.3 - The Relationship Between Power, β, and α Up 5.6 - Comparing Two Population Means ›