Hypothesis Test for Two Proportions

Ha Khanh Nguyen (hknguyen)

\[ X_1, X_2, \ldots, X_{n_1} \]
\[ X_1, X_2, \ldots, X_{n_1} \]
\[ \text{from population 1} \]
\[ \text{with “success” proportion } p_1 \]
\[ Y_1, Y_2, \ldots, Y_{n_2} \]
\[ Y_1, Y_2, \ldots, Y_{n_2} \]
\[ \text{from population 2} \]
\[ \text{with “success” proportion } p_2 \]

Consider the sample proportions

\[ \hat{p}_1 = \frac{x_1}{n_1} \quad \text{and} \quad \hat{p}_2 = \frac{x_2}{n_2} \]

where \( n_1 \) and \( n_2 \) are the sample sizes and \( x_1 \) and \( x_2 \) are the number of “successes” in the two samples from populations 1 and 2, respectively.

When all of the following conditions are satisfied,

\[ n_1 \cdot \hat{p}_1 \geq 10 \quad \text{and} \quad n_1 \cdot (1 - \hat{p}_1) \geq 10 \]
\[ n_2 \cdot \hat{p}_2 \geq 10 \quad \text{and} \quad n_2 \cdot (1 - \hat{p}_2) \geq 10 \]

then we can perform a hypothesis test for:

\[ H_0: p_1 = p_2 \quad \text{vs.} \quad H_1: p_1 < p_2 \]
\[ H_1: p_1 > p_2 \]
\[ H_1: p_1 \neq p_2 \quad \text{(2-sided test)} \]

Test statistic:

\[ Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p} \cdot (1 - \hat{p}) \cdot \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}} \]

where \( \hat{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{n_1 \cdot \hat{p}_1 + n_2 \cdot \hat{p}_2}{n_1 + n_2} \)

Example 1:
In a comparative study of two new drugs, A and B, 120 patients were treated with drug A and 150 patients with drug B, and the following results were obtained.

<table>
<thead>
<tr>
<th></th>
<th>Drug A</th>
<th>Drug B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cured</td>
<td>78</td>
<td>111</td>
</tr>
<tr>
<td>Not cured</td>
<td>42</td>
<td>39</td>
</tr>
<tr>
<td>Total</td>
<td>120</td>
<td>150</td>
</tr>
</tbody>
</table>

a) We wish to test whether drug B has a higher cure rate than drug A. Conduct the appropriate test at 5% significance level.

\[ p_B > p_A \]
Step 0: Check the assumptions

- If # of successes and failures in each sample ≥ 10
  - Test is ok to be applied

Step 1: State the hypotheses

- Ho: \( p_A = p_B \) vs. \( H_1: p_A < p_B \)

Step 2: Compute the test statistic

\[
Z = \frac{\hat{p}_A - \hat{p}_B}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \approx -1.6036
\]

Step 3: Compute the p-value

- \( H_1: p_A < p_B \) (left-side)

\[
p\text{-value} = P(Z < -1.6036) = 0.05440
\]

Step 4: State the conclusion

- If \( \alpha = 0.05 \), we fail to reject \( H_0 \). We conclude that the data does not provide sufficient evidence that the cure rate of drug B is higher than that of drug A.

Connection between 2-sided HT and 2-sided CI:

- \( H_0: p_A = p_B \) vs. \( H_1: p_A \neq p_B \)

\[
z = \frac{\hat{p}_A - \hat{p}_B}{\sqrt{\hat{p}(1-\hat{p}) \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}} \approx -1.6036 \text{ fail to reject } H_0
\]

- If \( p\text{-value} = 0.1088 > \alpha = 0.05 \), we fail to reject \( H_0 \).

\[
p\text{-value} \approx 0.1088 \Rightarrow \text{reject } H_0
\]