8.1 Comparing Two Population Proportions: Independent Sampling

Properties of the Sampling Distribution of \((\hat{p}_1 - \hat{p}_2)\)

1. The mean of the sampling distribution of \((\hat{p}_1 - \hat{p}_2)\) is \((p_1 - p_2)\).

\[
E(\hat{p}_1 - \hat{p}_2) = p_1 - p_2
\]

Thus, \((\hat{p}_1 - \hat{p}_2)\) is an unbiased estimator of \((p_1 - p_2)\).

2. The standard deviation of the sampling distribution of \((\hat{p}_1 - \hat{p}_2)\) is

\[
\sigma_{(\hat{p}_1 - \hat{p}_2)} = \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}
\]

3. If the sample sizes \(n_1\) and \(n_2\) are large, the sampling distribution of \((\hat{p}_1 - \hat{p}_2)\) is approximately normal.

**Large-Sample 100(1 - \(\alpha\))\% Confidence Interval for \((p_1 - p_2)\)**

\[
(\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2}\sigma_{(\hat{p}_1 - \hat{p}_2)} = (\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2} \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}} 
\]

\[
\approx (\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2} \sqrt{\frac{\hat{p}_1}{n_1} + \frac{\hat{p}_2}{n_2}}
\]

**Large-Sample Test of Hypothesis for \((p_1 - p_2)\)**

<table>
<thead>
<tr>
<th>One-Tailed Test</th>
<th>Two-Tailed Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>(H_0: (p_1 - p_2) = 0)</td>
<td>(H_0: (p_1 - p_2) = 0)</td>
</tr>
<tr>
<td>(H_a: (p_1 - p_2) &lt; 0)</td>
<td>(H_a: (p_1 - p_2) \neq 0)</td>
</tr>
<tr>
<td>[or (H_a: (p_1 - p_2) &gt; 0)]</td>
<td>[or (H_a: (p_1 - p_2) &gt; 0)]</td>
</tr>
</tbody>
</table>

Test statistic:

\[z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sigma_{(\hat{p}_1 - \hat{p}_2)}}\]

Rejection region: \(z < -z_\alpha\) or \(z > z_\alpha\) when \(H_a: (p_1 - p_2) > 0\)

Note: \(\sigma_{(\hat{p}_1 - \hat{p}_2)} = \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}} \approx \sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}\) where \(\hat{p} = \frac{x_1 + x_2}{n_1 + n_2}\)
Conditions Required for Valid Large-Sample Inferences about \((p_1 - p_2)\)

1. The two samples are randomly selected in an independent manner from the two target populations.

2. The sample sizes, \(n_1\) and \(n_2\), are both large, so the sampling distribution of \((\hat{p}_1 - \hat{p}_2)\) will be approximately normal. (This condition will be satisfied if both \(n_1\hat{p}_1 \geq 15\), \(n_1\hat{q}_1 \geq 15\), and \(n_2\hat{p}_2 \geq 15\), \(n_2\hat{q}_2 \geq 15\).)

Determination of Sample Size for Estimating \(p_1 - p_2\)

To estimate \((p_1 - p_2)\) to within a given sampling error \(SE\) and with confidence level \((1 - \alpha)\), use the following formula to solve for equal sample sizes that will achieve the desired reliability:

\[
n_1 = n_2 = \frac{(z_{\alpha/2})^2(p_1q_1 + p_2q_2)}{(SE)^2}
\]

You will need to substitute estimates for the values of \(p_1\) and \(p_2\) before solving for the sample size. These estimates might be based on prior samples, obtained from educated guesses or, most conservatively, specified as \(p_1 = p_2 = 0.5\).