Two sample binomials results

Recall $X \sim \text{Bin}(n_1, p_1)$ and $Y \sim \text{Bin}(n_2, p_2)$. Also this information is often arranged in a $2 \times 2$ table:

<table>
<thead>
<tr>
<th></th>
<th>$n_{11} = x$</th>
<th>$n_{12} = n_1 - x$</th>
<th>$n_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n_{21} = y$</td>
<td>$n_{22} = n_2 - y$</td>
<td>$n_2$</td>
<td></td>
</tr>
</tbody>
</table>

- $\hat{R}D = \hat{p}_1 - \hat{p}_2$

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

- $\hat{R}R = \frac{\hat{p}_1}{\hat{p}_2}$

$$SE_{\log \hat{R}R} = \sqrt{\frac{(1-\hat{p}_1)}{\hat{p}_1 n_1} + \frac{(1-\hat{p}_1)}{\hat{p}_1 n_1}}$$

- $\hat{O}R = \frac{\hat{p}_1/(1-\hat{p}_1)}{\hat{p}_2/(1-\hat{p}_2)} = \frac{n_{11}n_{22}}{n_{12}n_{21}}$

$$SE_{\log \hat{O}R} = \sqrt{\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}}$$

$CI = \text{Estimate} \pm Z_{1-\alpha/2} \cdot SE_{\text{Est}}$
Standard errors

- **delta method** can be used to obtain large sample standard errors

- Formally, the delta methods states that if

  \[
  \frac{\hat{\theta} - \theta}{\hat{SE}_\hat{\theta}} \rightarrow N(0, 1)
  \]

  then

  \[
  \frac{f(\hat{\theta}) - f(\theta)}{f'(\hat{\theta})\hat{SE}_\hat{\theta}} \rightarrow N(0, 1)
  \]

- Asymptotic mean of \( f(\hat{\theta}) \) is \( f(\theta) \)

- Asymptotic standard error of \( f(\hat{\theta}) \) can be estimated with \( f'(\hat{\theta})\hat{SE}_\hat{\theta} \)
Example

- $\theta = p_1$
- $\hat{\theta} = \hat{p}_1$
- $SE_{\hat{\theta}} = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}}$
- $f(x) = \log(x)$
- $f'(x) = 1/x$
- $\frac{\hat{\theta} - \theta}{SE_{\hat{\theta}}} \rightarrow N(0, 1)$ by the CLT

Then $SE_{\log \hat{p}_1} = f'(\hat{\theta})SE_{\hat{\theta}}$

$$= \frac{1}{\hat{p}_1} \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1}} = \sqrt{\frac{(1 - \hat{p}_1)}{\hat{p}_1 n_1}}$$

And

$$\frac{\log \hat{p}_1 - \log p_1}{\sqrt{\frac{(1-\hat{p}_1)}{\hat{p}_1 n_1}}} \rightarrow N(0, 1)$$
Putting it all together

- Asymptotic standard error

\[
\text{Var}(\log \hat{RR}) = \text{Var}\{\log(\hat{p}_1/\hat{p}_2)\} \\
= \text{Var}(\log \hat{p}_1) + \text{Var}(\log \hat{p}_2) \\
\approx \frac{(1 - \hat{p}_1)}{\hat{p}_1 n_1} + \frac{(1 - \hat{p}_2)}{\hat{p}_2 n_2}
\]

- The last line following from the delta method

- The approximation requires large sample sizes

- The delta method can be used similarly for the log odds ratio
Motivation for the delta method

- If $\hat{\theta}$ is close to $\theta$ then
  \[
  \frac{f(\hat{\theta}) - f(\theta)}{\hat{\theta} - \theta} \approx f'(\hat{\theta})
  \]

- So
  \[
  \frac{f(\hat{\theta}) - f(\theta)}{f'(\hat{\theta})} \approx \frac{\hat{\theta} - \theta}{SE_{\hat{\theta}}}
  \]

- Therefore
  \[
  \frac{f(\hat{\theta}) - f(\theta)}{f'(\hat{\theta})SE_{\hat{\theta}}} \approx \frac{\hat{\theta} - \theta}{SE_{\hat{\theta}}}
  \]