

# 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:

$n_{11} = x$	$n_{12} = n_1 - x$	$n_1$
$n_{21} = y$	$n_{22} = n_2 - y$	$n_2$

- $\hat{RD} = \hat{p}_1 - \hat{p}_2$

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

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

$$\hat{SE}_{\log \hat{RR}} = \sqrt{\frac{(1-\hat{p}_1)}{\hat{p}_1 n_1} + \frac{(1-\hat{p}_2)}{\hat{p}_2 n_2}}$$

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

$$\hat{SE}_{\log \hat{OR}} = \sqrt{\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}}$$

$$\text{CI} = \text{Estimate} \pm Z_{1-\alpha/2} \text{SE}_{\text{Est}}$$

# Standard errors

---

- **delta method** can be used to obtain large sample standard errors
- Formally, the delta method states that if

$$\frac{\hat{\theta} - \theta}{\hat{S}E_{\hat{\theta}}} \rightarrow N(0, 1)$$

then

$$\frac{f(\hat{\theta}) - f(\theta)}{f'(\hat{\theta})\hat{S}E_{\hat{\theta}}} \rightarrow N(0, 1)$$

- Asymptotic mean of  $f(\hat{\theta})$  is  $f(\theta)$
- Asymptotic standard error of  $f(\hat{\theta})$  can be estimated with  $f'(\hat{\theta})\hat{S}E_{\hat{\theta}}$

# Example

---

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

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

- 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

$$\begin{aligned}\text{Var}(\log \hat{RR}) &= \text{Var}\{\log(\hat{p}_1/\hat{p}_2)\} \\ &= \text{Var}(\log \hat{p}_1) + \text{Var}(\log \hat{p}_2) \\ &\approx \frac{(1 - \hat{p}_1)}{\hat{p}_1 n_1} + \frac{(1 - \hat{p}_2)}{\hat{p}_2 n_2}\end{aligned}$$

- 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

$$\frac{f(\hat{\theta}) - f(\theta)}{f'(\hat{\theta})} \approx \hat{\theta} - \theta$$

- Therefore

$$\frac{f(\hat{\theta}) - f(\theta)}{f'(\hat{\theta})\hat{SE}_{\hat{\theta}}} \approx \frac{\hat{\theta} - \theta}{\hat{SE}_{\hat{\theta}}}$$