Two Proportions

Parameter: $p_1 - p_2$

Point estimate: $\hat{p}_1 - \hat{p}_2$

Hypothesis Test

Hypotheses

$H_0: p_1 - p_2 = 0$

H_A:
① $p_1 - p_2 < 0$
② $p_1 - p_2 \neq 0$
③ $p_1 - p_2 > 0$

Test Statistic Random Variable (Assuming H_0)

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n_1} + \frac{\hat{p}(1-\hat{p})}{n_2}}} \sim N(0,1)$$

where

$$\hat{p} = \frac{\#\text{ successes}}{\#\text{ cases}} = \frac{\hat{p}_{1,obs}\, n_1 + \hat{p}_{2,obs}\, n_2}{n_1 + n_2}$$

Observed Test Statistic

$$z_{obs} = \frac{\hat{p}_{1,obs} - \hat{p}_{2,obs}}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n_1} + \frac{\hat{p}(1-\hat{p})}{n_2}}}$$

P-value

① $P(\hat{p}_1 - \hat{p}_2 \leq \hat{p}_{1,obs} - \hat{p}_{2,obs}) = P(Z \leq z_{obs})$

② $P(|\hat{p}_1 - \hat{p}_2| \geq |\hat{p}_{1,obs} - \hat{p}_{2,obs}|) = P(|Z| \geq |z_{obs}|)$

③ $P(\hat{p}_1 - \hat{p}_2 \geq \hat{p}_{1,obs} - \hat{p}_{2,obs}) = P(Z \geq z_{obs})$

Conditions for Distributional Approximation (Assuming H_0 is true)

1. Independent observations in each sample
2. Independent selection of samples
3. Number of pooled successes & pooled failures is at least 10
   [$n_1\hat{p} \geq 10$, $n_2\hat{p} \geq 10$, $n_1(1-\hat{p}) \geq 10$, $n_2(1-\hat{p}) \geq 10$]

Confidence Interval

Formula for CI

$$\hat{p}_{1,obs} - \hat{p}_{2,obs} \pm z^* \sqrt{\frac{\hat{p}_{1,obs}(1-\hat{p}_{1,obs})}{n_1} + \frac{\hat{p}_{2,obs}(1-\hat{p}_{2,obs})}{n_2}}$$

Conditions for Distributional Approximation

1. Independent observations in each sample
2. Independent selection of samples
3. Number of observed successes and failures is at least 10 for each group
   $n_1 \hat{p}_{1,obs} \geq 10$
   $n_1(1-\hat{p}_{1,obs}) \geq 10$
   $n_2 \hat{p}_{2,obs} \geq 10$
   $n_2(1-\hat{p}_{2,obs}) \geq 10$