Foundations of Data Science: A/B testing - Large sample theory

Large sample theory

$$Z = \frac{\hat{\vartheta} - \vartheta}{\hat{\sigma}_{\hat{\vartheta}}}$$

Parameter = difference between the population proportions

$$d = p_A - p_B$$

Estimator: difference between sample proportions:

$$\hat{d} = \hat{p}_A - \hat{p}_B \qquad \hat{p}_A = \frac{n_A}{n} \qquad \hat{p}_B = \frac{n_B}{n}$$

$$\hat{\sigma}^2_{\hat{d}} = \hat{\sigma}^2_{\hat{p}_A} + \hat{\sigma}^2_{\hat{p}_B} \qquad \hat{\sigma}^2_{\hat{p}_A} = \frac{\hat{p}_A(1-\hat{p}_A)}{n} \qquad \hat{\sigma}^2_{\hat{p}_B} = \frac{\hat{p}_B(1-\hat{p}_B)}{n}$$

$$(1-\alpha)\ CI = \left(\hat{d} - z_{\alpha/2}\,\hat{\sigma}_{\hat{d}}\ ,\ \hat{d} + z_{\alpha/2}\,\hat{\sigma}_{\hat{d}}\right)$$

Eg. $\hat{d} = \hat{p}_A - \hat{p}_B = 0.70 - 0.72 = -0.02$

$$\hat{\sigma}_{\hat{d}} = \frac{\sqrt{\hat{p}_A(1-\hat{p}_A) + \hat{p}_B(1-\hat{p}_B)}}{\sqrt{n}}$$

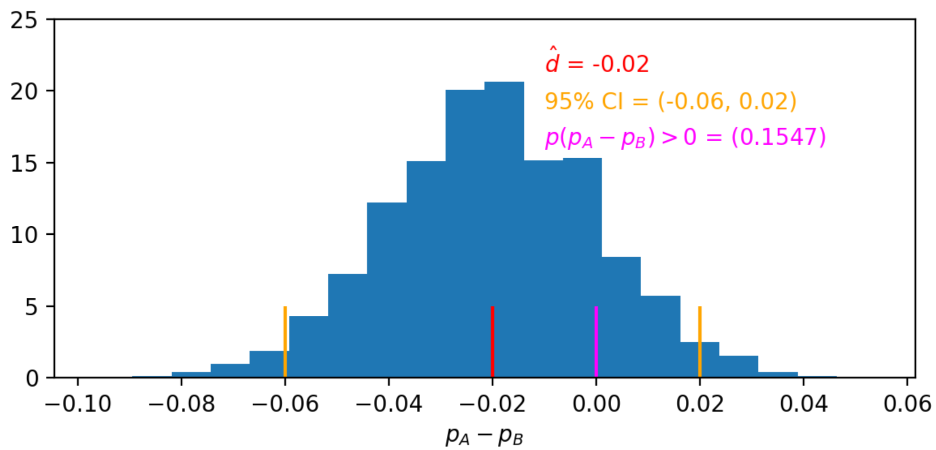
$$= \frac{\sqrt{0.70(1-0.70) + 0.72(1-0.72)}}{\sqrt{1000}} = 0.020$$

$$95\% \text{ CI} \Rightarrow z_{\alpha/2} = z_{0.025} = 1.96$$

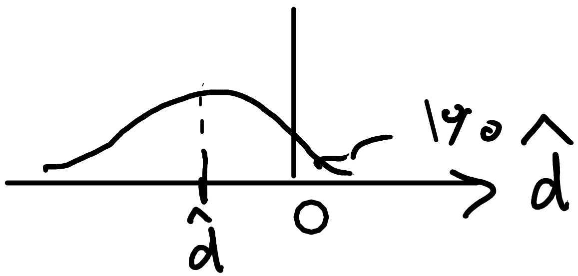
$$\Rightarrow \text{CI}: \left(\hat{d} - z_{\alpha/2}\,\hat{\sigma}_{\hat{d}}\ ,\ \hat{d} + z_{\alpha/2}\,\hat{\sigma}_{\hat{d}}\right)$$

$$= -0.02 - 1.96 \times 0.020,\ 0.02 + 1.96 \times 0.02$$

$$= \underline{\underline{(-0.06,\ 0.02)}}$$



Sample size calculation



$$\frac{0 - \hat{d}}{\hat{\sigma}_{\hat{d}}} = z_{0.99}$$

$$\hat{\sigma}_{\hat{d}} = \frac{\sqrt{\hat{p}_A(1-\hat{p}_A) + \hat{p}_B(1-\hat{p}_B)}}{\sqrt{n}}$$

$$\frac{-\sqrt{n}\,\hat{d}}{\sqrt{\hat{p}_A(1-\hat{p}_A) + \hat{p}_B(1-\hat{p}_B)}} = z_{0.99}$$

$$\Rightarrow \quad n = \frac{z_{0.99}^2}{\hat{d}^2}\left(\hat{p}_A(1-\hat{p}_A) + \hat{p}_B(1-\hat{p}_B)\right)$$