The background of the slide features a stylized globe on the left side, partially obscured by a dense field of binary code (0s and 1s) that recedes into the distance, creating a sense of depth and digital connectivity. The overall color palette is a mix of light blues, purples, and whites.

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

## Large sample theory

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

Parameter = difference between the population proportions

$$d = p_A - p_B$$

Estimator: difference between sample proportions:

$$\hat{d} = \hat{p}_A - \hat{p}_B$$

$$\hat{p}_A = \frac{n_A}{n}$$

$$\hat{p}_B = \frac{n_B}{n}$$

$$\hat{\sigma}_{\hat{d}}^2 = \hat{\sigma}_{\hat{p}_A}^2 + \hat{\sigma}_{\hat{p}_B}^2$$

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

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

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

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

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

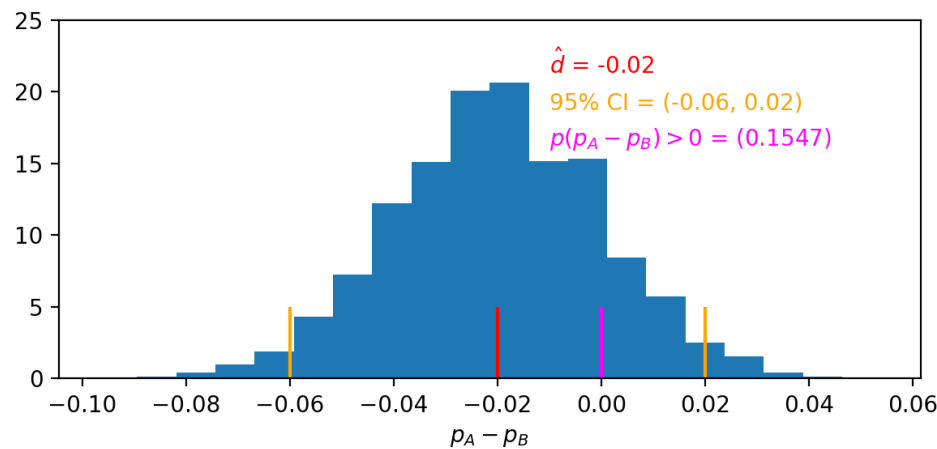
$$= \frac{\sqrt{0.70(1-0.70) + 0.72(1-0.72)}}{\sqrt{1000}} = 0.020$$

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

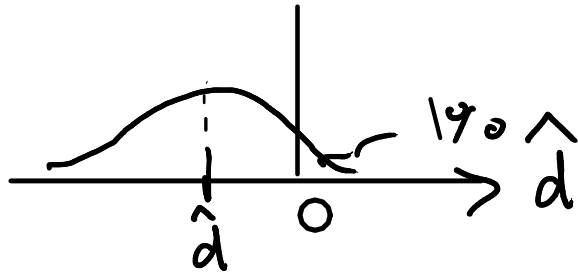
$$\Rightarrow \text{CI} : \left( \hat{d} - z_{\alpha/2} \hat{\sigma}_{\hat{d}}, \hat{d} + z_{\alpha/2} \hat{\sigma}_{\hat{d}} \right)$$

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

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



# Sample size calculation



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

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

$$-\sqrt{n} \hat{d}$$

$$= z_{0.99}$$

$$\sqrt{\hat{p}_A(1-\hat{p}_A) + \hat{p}_B(1-\hat{p}_B)}$$

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

---

---