ONLINE MASTERS IN **DATA SCIENCE**

DSC 215 - PROBABILITY AND STATISTICS FOR DATA SCIENCE

INFERENCE FOR A SINGLE PROPORTION

UC San Diego

COMPUTER SCIENCE & ENGINEERING

HALICIOĞLU DATA SCIENCE INSTITUTE

Confidence Interval for a Single Proportion

- **Recall: there are 4 steps to construct a confidence interval for a single proportion**
- **Step 1 (Prepare):** Identify $\hat{p}$ and n. Decide on what confidence interval to use.
- **Step 2 (Check):** Check the conditions for using the CLT (independence & success/ failure). **Use $\hat{p}$ instead of p** to check the Success/Failure Condition.
- **Step 3 (Calculate):** If the conditions hold, calculate SE (using $\hat{p}$), find $z^\star$, construct the interval.
- **Step 4 (Conclude):** Interpret your CI and provide your conclusions in the context of the problem at hand.

Choosing a Sample Size for Estimating $\hat{p}$

- **Remember:**

A confidence interval is given by

$$I = (\hat{p} - z^\star \cdot SE_{\hat{p}} \quad , \quad \hat{p} + z^\star \cdot SE_{\hat{p}})$$

(For example when working with a $95\ \%$ Confidence Interval: $z^\star = 1.96$.)

Here, the standard error is given by

$$SE = \sqrt{\frac{p(1-p)}{n}}$$

Choosing a Sample Size for Estimating $\hat{p}$

- Define **the margin of error:** $z^{\star} \cdot SE_{\hat{p}}$
- We often write $\hat{p} \pm (z^{\star} \cdot SE_{\hat{p}})$ to indicate that the confidence interval is $I = (\hat{p} - z^{\star} \cdot SE_{\hat{p}} \quad , \quad \hat{p} + z^{\star} \cdot SE_{\hat{p}})$
- **Example:** Suppose you want to find a sample size n so that the sample proportion is within ± 0.05 of the population proportion in a $95\,\%$ Confidence Interval.

$$95\,\%\ CI \implies z^{\star} = 1.96. \qquad SE = \sqrt{\frac{p(1-p)}{n}}$$

Want to solve for n, so that $z^{\star} \times \sqrt{\frac{p(1-p)}{n}} \leq 0.05$

Choosing a Sample Size for Estimating $\hat{p}$

- Want to solve for n, so that $z^\star \times \sqrt{\dfrac{p(1-p)}{n}} \leq 0.05$
- **Issue:** We may not know p! **Workaround:** *If we have a good estimate for p, use that.*
- *If not*, then **notice that requiring**

$$z^\star \times \sqrt{\frac{p(1-p)}{n}} \leq 0.05 \iff n \geq \frac{(z^\star)^2}{(0.05)^2}\, p(1-p)$$

- The right hand side is largest when $p = 0.5$, so it is enough to pick $n \geq \dfrac{(z^\star)^2}{(0.05)^2} 0.25$
- In our example $z^\star = 1.96$, so it's enough to pick $n \geq 1.96^2 \times \dfrac{0.25}{0.05^2} = 384.16$