M358K: November 2^{nd}, 2022.

Statistical Inference for a Single Proportion.

Let p denote our population parameter, i.e.,
the parameter p represents the probability that a randomly chosen member of the population has a particular trait
(e.g., they will vote for the purple party,
have a car or not, ...).

In other words, p stands for the probability of "success" in a single trail.

Plan: Use the sample proportion as a suitable statistic to study p from a well-designed sample.

Let n be the sample size.

Let X be the count random variable,

i.e., the # of times the particular trait of interest will be observed in the sample,

i.e., the # of "successes" in n independent trials w/ the probability of "success" in every trial equal to p.

=> The sampling distribution of the count random variable is:

$$\boxed{X \sim \text{Binomial}(\#\text{ of trials} = \text{sample size} = n, \text{ probab. of "success"} = p)}$$

Our unknown parameter of interest!

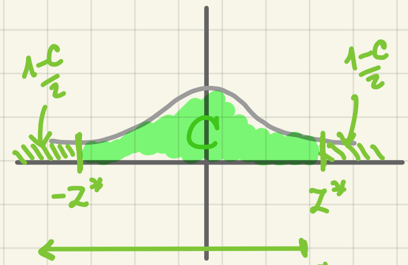
$\hat{p}$ ... the proportion of "successes" in our sample, i.e.,
the sample proportion, i.e., $\boxed{\hat{p} = \frac{X}{n}}$

For "large" sample sizes n, i.e., w/ $\boxed{n \cdot p > 10}$ and $\boxed{n(1-p) > 10}$,
we can use the normal approximation to the binomial, i.e.,

$$\boxed{\begin{aligned} X &\approx \boxed{\text{Normal}}\left(\text{mean} = n \cdot p,\ \text{sd} = \sqrt{n \cdot p \cdot (1-p)}\right) \\ \hat{p} &\approx \boxed{\text{Normal}}\left(\text{mean} = p,\ \text{sd} = \boxed{\sqrt{\frac{p(1-p)}{n}}}\right) \end{aligned}}$$

Confidence Intervals for p.

Let C be our confidence level.



$$C + \frac{1-C}{2} = \frac{1+C}{2}$$

pt. estimate $\pm$ margin-of error

$$z^* \cdot \underbrace{(\text{std error})}_{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$$

w/ $z^* = \Phi^{-1}\left(\frac{1+C}{2}\right)$

$= \text{qnorm}\left(\frac{1+C}{2}\right)$

$$\boxed{\hat{p} \pm z^* \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$$

$\hat{p}$: observed sample proportion

$$\boxed{\begin{aligned} &\text{If } n \cdot \hat{p} > 10 \\ &\text{and} \\ &n \cdot (1-\hat{p}) > 10 \end{aligned}}$$

Q: What is the smallest sample size necessary so that the margin of error is @ most a given value m?

→ : **We want:** $z^* \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq m$

Problem: We don't have $\hat{p}$!

Option One: Use a previous study's results.

Option Two: The conservative choice : instead of $\hat{p}$ use $\frac{1}{2}$

$$z^* \cdot \sqrt{\frac{\frac{1}{4}}{n}} \leq m \quad /^2$$

$$(z^*)^2 \cdot \frac{\frac{1}{4}}{n} \leq m^2$$

$$\boxed{\frac{(z^*)^2}{4m^2} \leq n}$$