M358K: November 1st, 2021.

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, or they have blood type A^-, ...).

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

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

Let n denote the sample size.

Let X be the count random variable,

i.e., the # of time the particular trait of interest will be obsewed in the sample,

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

=> The sampling distribution of the count random variable:

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

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/ $\boxed{np > 10}$ and $\boxed{n(1-p) > 10}$, we can use the normal approximation to the binomial, i.e.,

$$X \approx \text{Normal}\left(\text{mean} = n \cdot p, \ \text{sd} = \sqrt{n \cdot p(1-p)}\right)$$

$$\Rightarrow \hat{p} \approx \text{Normal}\left(\text{mean} = p, \ \boxed{\text{sd} = \sqrt{\frac{p(1-p)}{n}}}\right)$$

Confidence Intervals for p.

Let C be our confidence level.

$$\text{pt. estimate} \pm \underbrace{\text{margin of error}}_{z^* \cdot \underbrace{(\text{std error})}_{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}}$$

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

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

$\hat{p}$: observed sample proportion

If $n \cdot \hat{p} > 10$ and $n \cdot (1-\hat{p}) > 10$.

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

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

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

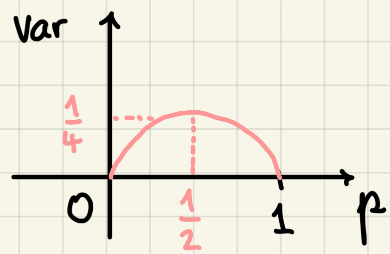
Option One: Use a previous study's results.

Option Two:

Q: What is the maximum variance that a single Bernoulli trial can have?

$$\text{Var}[\text{one trial}] = p(1-p)$$

Think about it as a f'tion of p.



The conservative choice for what to use instead of $\hat{p}$ is $1/2$.

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

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

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