M358K: October 27th, 2023.

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 our population has a particular trait

(e.g., they will vote for the orange party, wears glasses or not, ...).

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

Objective: Study the population proportion p.

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

Let n be our sample size.

Let X be the count random variable,

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

i.e., the # of "successes" in n independent trials where the probab. of "success" in every trial is equal to p.

$\Rightarrow$ The sampling dist'n of the count random variable is:

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

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., $\boxed{n \cdot p \geq 10}$ and $\boxed{n(1-p) \geq 10}$

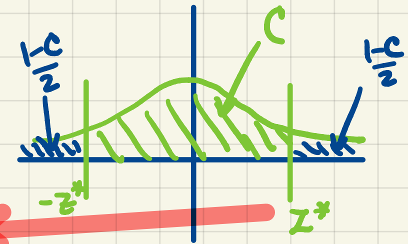
We can use the normal approximation to the binomial, i.e.,

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

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

Confidence Intervals for p.

Let C be our confidence level.



pt. estimate $\pm$ $\underbrace{\text{margin of error}}_{z^* (\text{std error})}$

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

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

std error $\rightarrow \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

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

$\hat{p}$: observed sample proportion

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

UNIVERSITY OF TEXAS AT AUSTIN

Problem Set # 15

Confidence intervals: One-sample proportion.

Problem 15.1. A simple random sample of 100 bags of tortilla chips produced by Company X is selected every hour for quality control. In the current sample, 18 bags had more chips (measured in weight) than the labeled quantity. The quality control inspector wishes to use this information to calculate a 90% confidence interval for the true proportion of bags of tortilla chips that contain more than the label states. What is the value of the standard error of $\hat{p}$?

$\rightarrow$: $\hat{p} = \frac{18}{100} = 0.18 \Rightarrow$ std error. $= \sqrt{\frac{0.18 \cdot 0.82}{100}} = \underline{0.038418}$

Problem 15.2. A simple random sample of 60 blood donors is taken to estimate the proportion of donors with type A blood with a 95% confidence interval. In the sample, there are 10 people with type A blood. What is the margin of error for this confidence interval?

$\rightarrow$: $\hat{p} = \frac{10}{60} \Rightarrow$ std error $= \sqrt{\frac{(\frac{1}{6})(\frac{5}{6})}{60}} = \underline{0.048}$

$\boxed{z^* = 1.96} \longrightarrow \times \Rightarrow$ margin of error $= \underline{0.0943}$

Problem 15.3. A simple random sample of 85 students is taken from a large university on the West Coast to estimate the proportion of students whose parents bought a car for them when they left for college. When interviewed, 51 students in the sample responded that their parents bought them a car. What is a 95% confidence interval for p, the population proportion of students whose parents bought a car for them when they left for college?

$\rightarrow$: $\hat{p} = \frac{51}{85} = 0.6 \Rightarrow$ std error $= \sqrt{\frac{0.6 \cdot 0.4}{85}} = \underline{0.05\ldots}$

$z^* = 1.96 \longrightarrow \times \Rightarrow$ margin of error $= \underline{0.104}$

Our confidence interval is:

$p = \boxed{0.6 \pm 0.1041}$

Problem 15.4. A simple random sample of 450 residents in the state of New York is taken to estimate the proportion of people who live within 1 mile of a hazardous waste site. It was found that $\boxed{135}$ of the residents in the sample live within 1 mile of a hazardous waste site.

(1) What are the values of the sample proportion of people who live within 1 mile of a hazardous waste site and its standard error?

$$\hat{p} = \frac{135}{450} = 0.3 \Rightarrow \text{std error} = \sqrt{\frac{0.3 \cdot 0.7}{450}} = \underline{0.0216}$$

(2) What are the values of the sample proportion of people who live outside of the 1 mile radius around a hazardous waste site and its standard error?

$$\hat{p}' = 0.7 \Rightarrow \text{std error} = \sqrt{\frac{0.7 \cdot 0.3}{450}} = \underline{0.0216}$$

(3) Do you notice something interesting about the above?

Problem 15.5. <u>**Sample size**</u>

The *Information Technology Department* at a large university wishes to estimate the proportion p of students living in the dormitories who own a computer. They want to costruct a 90% confidence interval. What is the minimum required sample size the IT Department should use to estimate the proportion p with a margin of error no larger than 2 percentage points?

Problem 15.6. (5 points)

You want to design a study to estimate the proportion of people who strongly oppose to have a state lottery. You will use a 99% confidence interval and you would like the margin of error of the interval to be 0.05 or less. What is the minimal sample size required?

a. 666
b. 543
c. 385
d. Not enough information is provided.
e. None of the above.