M358K: October 30th, 2023.

Single Proportion: Sample Size.

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

$\longrightarrow$:

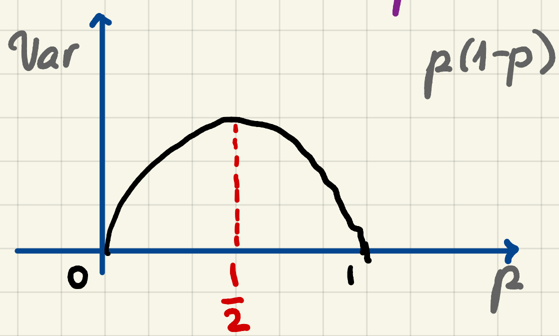
We want: $$z^* \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq m$$

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

Option One: Use the results of a previous study.

Option Two: The consewative choice

instead of $\hat{p}$ use $\frac{1}{2}$



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

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

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

$\square$

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 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?

(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?

(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?

$\rightarrow$:

$$n \geq \frac{(z^*)^2}{4m^2} = \frac{(1.645)^2}{4(0.02)^2} \approx \begin{cases} 1690.965 \\ \text{or} \\ 1691.266 \end{cases}$$

qnorm

$$n \geq 1691 \quad \text{or} \quad n \geq 1692$$

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?

tables $\rightarrow$ a. 666

b. 543

c. 385

d. Not enough information is provided.

e. None of the above.

$$n \geq \frac{(z^*)^2}{4m^2} = \frac{(\text{qnorm}(1.99/2))^2}{4(0.05)^2} = 663.4897$$

$$n \geq 664$$

Hypothesis Testing for the Population Proportion p

We test: $\boxed{H_0: p = p_0}$ vs. $H_a: \begin{cases} p < p_0 \\ p \neq p_0 \\ p > p_0 \end{cases}$

Test Statistic = ?

Sample Proportion of "successes" $\boxed{\hat{P}}$

We know its (approximate) sampling dist'n under the null hypothesis.

- For a large enough sample size, we know that:
  - the count r.v. satisfies:

    $$X \approx Normal\left(mean = n \cdot p_0,\ sd = \sqrt{n \cdot p_0(1-p_0)}\right)$$

  - the sample proportion r.v. satisfies:

$$\hat{P} = \frac{X}{n} \approx Normal\left(mean = p_0,\ sd = \sqrt{\frac{p_0(1-p_0)}{n}}\right)$$

The observed value of the sample proportion is denoted by $\hat{p}$.

The corresponding value of the z-statistic under the null hypothesis:

$$\boxed{z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}}$$

By def'n, the p-value is the probability of observing this z-score or something more extreme under the null.

IF $H_a: p < p_0$, then p-value $= \mathbb{P}[Z < z]$ = pnorm(z)

IF $H_a: p \neq p_0$, then p-value $= \mathbb{P}[Z < -|z|] + \mathbb{P}[Z > |z|]$

= 2*pnorm(-abs(z))

IF $H_a: p > p_0$, then p-value $= \mathbb{P}[Z > z]$ = 1-pnorm(z)

If a significance level α is given, then ....

IF p-value $\leq \alpha$, then REJECT THE NULL.

IF p-value $> \alpha$, then FAIL TO REJECT THE NULL.

UNIVERSITY OF TEXAS AT AUSTIN

Problem Set # 16

Hypothesis testing: One-sample proportion.

Problem 16.1. *Source: Problem* **8.99** *from the Moore/McCabe/Craig.*
Castaneda v. Partida is an important court case in which statistical methods were used as part of a legal argument. When reviewing this case, the Supreme Court used the phrase "two or three standard deviations" as a criterion for statistical significance. This Supreme Court review has served as the basis for many subsequent applications of statistical methods in legal settings. (The two or three standard deviations referred to by the Court are values of the z statistic and correspond to p-values of approximately 0.05 and 0.0026.)

In Castaneda the plaintiffs alleged that the method for selecting juries in a county in Texas was biased against Mexican Americans. For the period of time at issue, there were $181,535$ persons eligible for jury duty, of whom $143,611$ were Mexican Americans. Of the 870 people selected for jury duty, 339 were Mexican Americans.

(i) (1 point) What proportion of eligible jurors were Mexican Americans?

$$p_0 = \frac{143611}{181535} = 0.7911$$

(ii) (2 points) Let p denote the probability that a randomly selected juror is a Mexican American. Formulate the null and alternative hypotheses to be tested.

$$H_0: p = p_0 = 0.7911 \quad \text{vs.} \quad H_a: \underline{p < p_0} = 0.7911$$

(iii) (1 point) What is the sample proportion of jurors who were Mexican American?

$$\hat{p} = \frac{339}{870} = 0.3897$$

(iv) (4 points) Compute the $z-$statistic, and find the $p-$value.

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.3897 - 0.7911}{\sqrt{\frac{0.7911(1-0.7911)}{870}}} = -29.12641$$

$\Rightarrow$ p-value is virtually zero.

(v) (2 points) How would you summarize your conclusions? (A finding of statistical significance in this circumstance does not constitute proof of discrimination. It can be used, however, to establish a prima facie case. The burden of proof then shifts to the defense.)

There is evidence in favor of the prima facie case!