8.2 Estimating a Population Proportion

Conditions for estimating

1. Random: SRS (Inference + chance behavior) + 10% Condition
2. Large counts condition to show us we can use the Normal distribution

$n\hat{p} \geq 10 \quad n(1-\hat{p}) \geq 10$

# of success ←

$\cancel{n} \cdot \frac{X}{\cancel{n}} \geq 10$

# of success is more than 10

# of failures is more than 10

Ex: A quality control inspector takes a random sample of 25 bags of potato chips from the thousands of bags filled in an hour. Of the bags selected, 3 had too much salt. Are the conditions for estimating p met?

- Yes, there is a SRS chosen of 25 bags,
- 10% condition is met, since there is more than 250 bags of chips
- Large Counts not met, there are 3 bags that are too salty (failure) which is less than 10.

Constructing a CI for p

Inference about a population proportion p is based on a sampling $\hat{p}$. Therefore, as long as the 10% condition is met, we can use:

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$, since we don't know p, we

use: standard error of $\hat{p}$

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

Standard error: result of the standard deviation of statistic when estimated from data.

The <u>critical value</u>, z^*, is the standardized normal value needed to catch the central C% of area under the standard normal curve, assuming the large counts condition is met.

Ex: Use table A <u>and</u> your graphing calculator to find z^* for a 96% CI, assuming large counts is met.

$z^* = 2.05$

One sample z interval for a population proportion

when appropriate conditions are met, a C% CI for unknown proportion p is:

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

where z^* is the critical value for the standard normal curve with C% of its area between $-z^*$ & z^*

Alcohol abuse has been described by college presidents as the number one problem on campus, and it is an important cause of death in young adults. How common is it? A survey of 10,904 randomly selected US college students collected information on drinking behavior and alcohol related problems. The researchers defined "frequent binge drinking" as having 5 or more drinks in a row three or more times in the past two weeks. According to this definition, 2486 students were classified as frequent binge drinkers.

1. Identify the parameter of interest.

p = actual proportion US college students who are binge drinkers

2. Check the conditions for constructing a CI for the parameter.

Random: a random sample was chosen

10% condition: 10,904 is less than 10% of all US college students

Large counts: 2486 students binge drink (success) ≥ 10

8418 students don't binge drink (failure) ≥ 10

3. Find the z* for a 99% CI. Show your method. Then calculate the interval.

$z^* = 2.58$

$\hat{p} = \frac{2486}{10904} = .228$

$.228 \pm 2.58\left(\sqrt{\frac{.228(1-.228)}{10904}}\right)$

(.2176, .238)

4. Interpret the interval in context.

We are 99% confident that the interval from .2176 to .238 captures the actual proportion of US college students who binge drink

4 step process for calculating CI's

1. State: what parameter are we estimating and the confidence level.
2. Plan: Identify the appropriate inference method and check conditions
3. Do: Perform calculations if conditions are met.
4. Conclude: Interpret results in context.

Ex: Joe learned 70% of the Earth's surface is covered in water. He wondered if it was true so he had his dad toss an inflatable globe to him 50 times. When he caught the globe, he recorded where his left index finger was pointing. In 50 tosses, he pointed to water 33 times. Construct and interpret a 95% CI for the proportion of Earth's surface covered in water.

1. p = actual proportion of the Earth's surface covered in water with 95% confidence.

2. One proportion z interval for pop. proportion

Random: tossing a globe gives us a random sample

10% condition is met since all globe tosses are already independent

large counts: 33 success (water) ≥ 10

17 failures (land) ≥ 10

3. $.66 \pm 1.96\left(\sqrt{\frac{.66(.34)}{50}}\right)$

$(.53, .79)$

4. We are 95% confident that the interval from .53 to .79 capture the actual proportion of the Earth's surface covered in water.

Choosing a sample size for a desired margin of error.

To determine the sample size, n, that yields a C% CI for p with a maximum margin of error, ME, solve following for n:

$$z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < ME$$

The margin of error will always be less than or equal to ME if $\hat{p} = .5$