Confidence Intervals for Population Proportions

April 6th, 2020

Objective

By the end of this lesson, You should be able to…

1. Be able to describe the sampling distribution being used
2. Know the assumptions being made to conduct a 1 proportion confidence interval
3. Apply the 1 proportion confidence interval to a problem

Review Problem #1

Which of the following are true:

1. The sampling distribution of p-hat has a mean equal to the population proportion p
2. The sampling distribution of p-hat has a standard deviation equal to $\sqrt{(np(1-p))}$
3. The sampling distribution of p-hat is considered close to normal provided that n ≥ 30

Review #2

Which of the following are true statements?

1. The histogram of a binomial distribution with p=0.5 is always symmetric no matter what the value of n, the number of trials.
2. The histogram of a binomial distribution with p = 0.2 is skewed to the left.
3. The histogram of a binomial distribution with p = 0.9 looks more and more symmetric, the larger the value of n

Answers

Review #1: Only number 1 is true. The standard deviation of of p-hat is $\sqrt{\frac{p(1-p)}{n}}$

The sampling distribution is considered close to normal only if np and n(1-p) are greater than 10

Review #2: 1 and 3 are true statements. With a p = 0.2, there is a low probability of a lot of successes. There is a greater probability of a low number of successes. So the data piles up on the left side of the graph, creating a skewed right distribution.

Sampling distribution for a 1 prop Z interval

- Center: The mean of all sample proportions is equal to the population proportion. This informs us that the sampling distribution is an unbiased estimator of the population proportion as p-hat is on average the same of the population proportion.

- Spread: The standard deviation of the p-hat sampling distribution is $\sqrt{\frac{p(1-p)}{n}}$

- Shape: When np and n(1-p) are greater than 10, we can assume the sampling distribution is normal. This is nice because it allows us to use a normal distribution.

Assumption for one proportion confidence intervals

In order to apply the one proportion z confidence interval, we need to make sure we meet three assumptions:

1. Random: The sample is from a random process, or the sample can fairly be considered representative
2. Independent: The process of sampling does not change the probability of the event happening, and if we are dealing with a finite population, that we sample less than 10% of the population.
3. Normal: We meet the np and n(1-p) greater than or equal to 10 condition.

Example problem

A husband and wife, Mike and Lori, share a digital music player that has a feature that randomly selects which song to play. A total of 2,384 songs were loaded onto the player, some by Mike and the rest by Lori. Suppose that when the player was in the random-selection mode, 13 of the first 50 songs selected were songs loaded by Lori. Construct and interpret a 90 percent confidence interval for the proportion of songs on the player that were loaded by Lori.

We will use the 4 step process outlined in the textbook to solve this problem.

State

We wish to estimate the proportion of songs that Lori loaded onto the digital music player.

P: the true proportion of songs that Lori loaded onto the digital music player

Plan

We will use a 1 proportion z confidence interval with 90% confidence level

Random - Since the 50 songs are selected by the random-selection mode of the digital player, we will assume this represents a random selection of 50 songs from the music library.

Independent - We must assume that one song playing does not change the probability of another song playing afterward. Since it does not indicate replacement, we will hazard on the side of caution and assume no replacement. However, since 50 songs is less than 10% of the 2,384 songs in the library, we should be safe to assume normality.

Normal: p-hat = 0.26. np = 50*0.26 = 13, n(1-p) = 50*(1-0.26) = 37. Since both values are greater than 10, we can assume that a the sampling distribution is approximately normal.

Do

x= 13

n = 50

p-hat = 0.26

Confidence level = 90%

Critical value = 1.645 ← pulled from the bottom row of the t-table

Formula:

$$C.I. = \bar{p} \pm z^*\left(\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}\right) = 0.26 \pm 1.645\left(\sqrt{\frac{0.26(1-0.26)}{50}}\right) = 0.26 \pm 0.0620 = (0.1980,\ 0.3220)$$

Conclude

We are 90% confident, that the interval 0.1980 to 0.3220 captures the true population proportion of songs that Lori loaded onto the digital music player.

Common mistakes

1. Make sure to restate the problem that you are trying to solve.
2. When addressing the assumptions, do not just plug the numbers in. Also state why it meets or fails to meet the assumption. This is as much about communication as it is about the calculations.
3. Be very careful stating your conclusion. Do not use any language that implies the population proportion is moving around. It is a constant that you are trying to estimate. The sample proportion and confidence interval are the values that move around.

You try

A simple random sample of 1100 males aged 12 to 17 in the United States were asked whether they played massive multiplayer online role-playing games (MMORPGs); 775 said that they did. We want to use this information to construct a 95% confidence interval to estimate the proportion of all U.S. males aged 12 to 17 who play MMORPGs.

Answers - state

State: We wish to estimate the true proportion of U.S. males between the age of 12 and 17 that play MMORPG

P: true proportion of U.S. males aged 12 to 17 in the U.S. who play MMORPG's.

Answers - Plan

Random: The problem states that an SRS was taken.

Normal: np = 77, n(1-p)= 325, Since both np and n(1-p) are greater than 10, we can fairly assume normality.

Independence: It seems reasonable to assume that individual observations are independent, and the population of U.S. males aged 12 to 17 is certainly more than 10 times the sample size.

Answers - Do

$$\sqrt{\frac{(0.705)(0.295)}{1100}} \approx 0.0253$$

$$0.705 \pm 1.96(0.0253) \rightarrow 0.705 \pm 0.050 \rightarrow (0.655, 0.755)$$

Answers - Conclude

We are 95% confident that the interval from 0.655 to 0.755 contains the true proportion of males

aged 12 to 17 in the U.S. who play MMORPG's

Extra examples and practice

Reading: pg 484-494

HW: 35, 37, 41, 43, 47