**MA206**, Lesson 10 - Size of Effect

**Review**: How do we calculate the standardized statistic for one mean?

$t =$

**Review**: How do we calculate the confidence interval for one **proportion**?

How do we calculate the confidence interval for one **mean**?

How do we interpret a (95%) confidence interval?

What three factors impact the width of our confidence interval?

What are some cautions when conducting inference?

**1)** Suppose we are constructing a confidence interval using simulation. Using two-sided hypothesis tests each time with the following null hypothesized values, we obtain the resulting p-values.

| **Null Hypothesis** | p-value | | Null Hypothesis | p-value |
|---|---|---|---|---|
| 0.45 | 0.007 | | 0.53 | 0.602 |
| 0.46 | 0.012 | | 0.54 | 0.124 |
| 0.47 | 0.045 | | 0.55 | 0.084 |
| 0.48 | 0.079 | | 0.56 | 0.052 |
| 0.49 | 0.121 | | 0.57 | 0.034 |
| 0.50 | 0.254 | | 0.58 | 0.019 |
| 0.51 | 0.643 | | 0.59 | 0.012 |
| 0.52 | 0.986 | | 0.60 | 0.004 |

**a)** What is the 95% Confidence Interval, given the above results?

**b)** What is the 99% Confidence Interval, given the above results?

**2)** According to a 2018 report by the U.S. Department of Labor, civilian Americans spend 2.84 hours per day watching television. A faculty researcher, Dr. Sameer, at California Polytechnic State University (Cal Poly) conducts a study to see whether a different average applies to Cal Poly students. Suppose that for a random sample of 100 Cal Poly students, the mean and standard deviation of hours per day spent watching TV turns out to be 3.01 and 1.97 hours, respectively. There is not strong skew.

**a)** Is our statistic quantitative or categorical?

**b)** What is the value of our statistic (hint: $\hat{p}$ *and/or* $\bar{x}$ *and/or* $s$)?

**c)** Do we meet our validity conditions?

**d)** What is our 95% Confidence Interval for the true mean hours that Cal Poly students spend watching television per day?

**e)** Given our confidence interval above, what do we know about the results of a strength of evidence test with a null hypothesis of $\mu = 2.84$ and an alternate hypothesis of $\mu \neq 2.84$?

**f)** Report your standardized statistic (t or z) and p-value given the above data and a null hypothesis of $\mu = 2.84$ and an alternate hypothesis of $\mu \neq 2.84$.

**3)** According to the 2019 National Coffee Drinking Study from the National Coffee Association, comprised of a survey sent to shops and customers registered with the NCA, a total of 1,774 of 2,815 U.S. adults respondents reported drinking coffee in the past 24 hours.

**a)** Is our statistic quantitative or categorical?

**b)** What is the value of our statistic (hint: $\hat{p}$ *and/or* $\bar{x}$ *and/or* $s$)?

**c)** Do we meet our validity conditions?

**d)** What is our 99% Confidence Interval for the true proportion of American coffee drinkers, given the data?

**e)** Report your standardized statistic (t or z) and p-value, given the above data and a null hypothesis that the true proportion of Americans who drink coffee is 70% with an alternate hypothesis that it is not 70%.

**f)** How broadly can you generalize these results?

**4)** The file AgeFirstChild.csv on Teams has data from a 2018 General Social Survey (GSS), which uses a random sample of U.S. adults. One question asked was the age of the respondent when their first child was born. Use the dataset to answer the question, what is the average age of women when they have their first liveborn child?

**a)** Do we meet validity conditions to use theoretical methods?

**b)** Why do we check for validity conditions?

**c)** Calculate a 99% confidence interval for the average age a mother gives birth to her first child.

**d)** Interpret this confidence interval.

**e)** Your friend states that they heard the average age for a first baby is 23 years old. Do you consider this feasible?

**5)** A recent survey by Pew Research (2009) ran a survey of 242 random cell phone users across the United States, ages 16 and 17, and asked if they had ever talked on a cell phone while driving. The results are available in the csv file "Phones.csv". Read the file into R and answer the following questions.

**a)** Generate a lot (bar plot) of the data, visualizing the proportions of those who answered "yes" and "no." (Hint: Sample code can be found on the course guide, Lesson 4-6).

**b)** Given the data, what is the 95% confidence interval for the true proportion of all cell phone users, aged 16 and 17, that have talked with a cell phone while driving?

**c)** Suppose an article is published estimating that up to 58% of Americans have talked on their cell phones while driving. Is this a valid claim?

**d)** Given the data, what is the 95% confidence interval for the true proportion of cell phone users **for only females**?

**e)** Given the data, what is the 95% confidence interval for the true proportion of cell phone users **for only males**??

**f)** Can we conclude a difference between male and female proportions of 16- and 17-year olds who have talked on the phone while driving?