MA206, Lesson 10 - Size of Effect

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

$t = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}}$

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

$\hat{p} \pm qnorm(1-\frac{\alpha}{2}) \times \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ alternatively, $\hat{p} \pm M \times SD(\hat{p})$

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

$\bar{x} \pm qt(1-\frac{\alpha}{2}, n-1) \times \frac{s}{\sqrt{n}}$ alternatively, $\bar{x} \pm M \times SD(\bar{x})$

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

"We are (95%) confidence that the true parameter lies within the range (_low_, _high_)

What three factors impact the width of our confidence interval?

- Confidence Level α
- Sample Size n
- Standard Deviation / Variation in our Data (quantitative only)

What are some cautions when conducting inference?

- **Nonrandom errors** This includes human error, sampling bias, etc.
- **Random errors** This includes random variability in the process
- **Statistical vs. Practical significance** Is the impact important in real-world practice

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?

(0.48, 0.56)

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

(0.46, 0.59)

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?

Quantitative

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

$\bar{x} = 3.01$, s=1.97

c) Do we meet our validity conditions?

Yes, we have at least 20 observations ($100 \geq 20$) and the data is not strongly skewed.

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

Confidence Interval $= \bar{x} \pm qt(1 - \frac{\alpha}{2}, n-1) \times \frac{s}{\sqrt{n}} \quad = \quad 3.01 \pm qt(1 - \frac{.05}{2}\ ,\ 99) \times \frac{1.97}{\sqrt{100}}$

$= (2.6191, 3.4009)$

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 $\mu \neq 2.84$?

We know that the p-value will be greater than 0.05, as 2.84 did "make the cut" and falls within our 95% confidence interval.

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 $\mu \neq 2.84$.

$t = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}} \quad = \quad \frac{3.01-2.84}{\frac{1.97}{\sqrt{100}}} \quad = \quad 0.8629$

p-value = 2*(1 - pt(abs(t), n-1)) = 2 * (1 - pt(abs(0.8629442), 99)) = 0.3903

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?

Categorical

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

$\hat{p} = \frac{1774}{2815} = 0.6302$

c) Do we meet our validity conditions?

Yes, for categorical we need at least 10 successes and 10 failures. Here, we have 1,774 and 1,041, both larger than 10.

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

$$\text{Confidence Interval} = \hat{p} \pm qnorm(1 - \tfrac{\alpha}{2}) \times \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \quad = \quad 0.63 \pm qnorm(1 - \tfrac{.01}{2}) \times \sqrt{\frac{0.63(1-63)}{2815}}$$
$$= (0.6068\ ,\ 0.6536)$$

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%.

$z = \frac{\hat{p}-\pi}{\sqrt{\frac{\pi(1-\pi)}{n}}} \quad = \quad \frac{0.63-0.7}{\sqrt{\frac{0.7(1-0.7)}{2815}}} \quad = \quad$ -8.0819

p-value = 2*(1 - pnorm(abs(z))) = 2 * 1 - pnorm(abs(-8.104525)) = 6.661338e^{-16}

f) How broadly can you generalize these results?

Assuming that the shops and customers surveyed was done randomly, these results could be generalized to shops and customers registered with the NCA, but could not be safely generalized to all Americans as there may be inherent bias in the sampling.

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?

Yes, we do. We have at least 20 observations and the data is not strongly skewed.

b) Why do we check for validity conditions?

Validity conditions allow us to use theoretical methods to calculate by using a normal or t distribution. The Central Limit Theorem states that, if sample size is large enough, the distribution of sample statistics will converge to be approximately normal. Validity conditions tell us our approximation will be a good fit.

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

(23.93735, 24.66265)

d) Interpret this confidence interval.

We are 99% confident that the true mean age that a woman has her first baby in the United States is between 23.937 and 24.663 years old.

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

No, this falls outside of our confidence interval, so we would not consider 23 a feasible age for the true average age of having a first baby.

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?

(0.4577, 0.5836)

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?

No, these results cannot be generalized to the entire American population. At best we could generalize to 16- and 17-year olds with cell phones, assuming the survey was conducting using random sampling.

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

(0.4808, 0.6531)

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

(0.3784, 0.5608)

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?

No, we cannot. Both 95% confidence intervals overlap, indicating that there are values which are plausible for both male and female young drivers. Therefore, it is plausible that they have the same proportions and there is no difference.