

MA 207 – Spring 2018
Lab #6: Hypothesis Tests and Confidence Intervals

- Your work should be compiled in a Word document and turned in via Moodle.
- Please write all of your answers in complete sentences using appropriate grammar, punctuation, and terminology.
- You should work in groups of two. You must both work on all problems.
- You will probably not be able to complete it within class time.
- Both students should submit the Word document, with both names at the top.

Refer to the *Instructions for Lab Write-ups* document on Moodle for information on how your lab should be formatted.

Problem 1:

In September 2016, the Gallup poll conducted a telephone interview to judge the opinions of the American public on the need for a third political party. The survey was conducted via telephone on adults, aged 18 and older, living in the United States. During the telephone survey they asked the following question.

In your view, do the Republican and Democratic parties do an adequate job of representing the American people, or do they do such a poor job that a third major party is needed?

If you would like to read the full article you can find it here: http://www.gallup.com/poll/195920/americans-desire-third-party-persists-election-year.aspx?g_source=ELECTION_2016&g_medium=topic&g_campaign=tiles

- A. Before running a survey it is good practice to determine the sample size that you need ahead of time. Instead of just picking a sample size out of the blue, it is typical to decide on a targeted margin of error ahead of time and then to work backwards to the sample size. Let's assume that Gallup wanted the margin of error to be less than 5%. Recall that the margin of error for a 95% confidence interval on a single proportion is:

$$MOE = 1.96 \times \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}$$

If we know that the margin of error is 5% and we make the conservative estimate that $\hat{p} = .5$ (50% are in favor of a third party), then what sample size is necessary? *Hint: Do some basic algebra to solve for the sample size.*

- B. The Gallup poll in the article surveyed 1,020 people. Using the same conservative estimate for the proportion of respondents who stated that a third party was necessary ($\hat{p} = .5$), approximate the margin of error for the Gallup survey. Is this approximation an over- or under-estimate of the true margin of error?
- C. The title of the article in question was “Americans? Desire for Third Party Persists This Election Year”, and the first sentence states that 57% of Americans say that a third major U.S. political party is needed. The “57%” is a point estimate for the actual feelings of the entire United States.
- What is the standard error of this estimate?
 - What is the actual margin of error for this proportion?
 - Create a 95% confidence interval around the 57% statistic that gives a range estimate for the proportion of US adults that think that a third major U.S. political party is needed. Then write a sentence explaining the meaning of “95% confidence” in this context.

- D. The article states that in the past election years (2008 and 2012) “...Americans were evenly divided on the need for a third party.” We will perform a hypothesis test on this to determine if the proportion has really changed or if we are just seeing natural variation in the random sample.
- The null hypothesis for this test is: We assume that the proportion of American adults that express a need for a third party is 0.5 (50%). What is the alternative hypothesis? Please state this in both proper symbols and in a sentence.
 - What is the standard error for the hypothesis test? (Be sure to base your standard error on the null hypothesis)
 - Calculate a test statistic for the 2016 proportion 0.57. That is, determine how many standard errors the 2016 value is above the null hypothesis.
 - Calculate the p-value for this hypothesis test.
 - If possible, draw a conclusion from this hypothesis test. Explain what your calculations mean in this particular context.

Problem 2:

Immigration has been a major topic in the news recently. A recent Pew report <http://www.pewhispanic.org/2015/09/28/chapter-4-u-s-public-has-mixed-views-of-immigrants-and-immigration/> looked at American attitudes toward immigrants. As you may expect, attitudes varied between subsections of the American population. Is it possible that these differences were due to sampling variability, or do they indicate statistically significant differences between subpopulations? The study asked Americans whether there were immigrants in their communities and then also asked them about whether immigrants make American society better in the long run. While the Pew report did not give exact numbers for how many people answered each question, the following chart provides an estimate.

Do you have immigrants living in your community?			
Yes “In community” group $n_1 = 1100$		No “Not in community” group $n_2 = 380$	
Do immigrants make American society better in the long run?		Do immigrants make American society better in the long run?	
Yes: 495	No: 605	Yes: 126	No: 254

The null hypothesis is: We assume that the true proportion of American adults who think that immigrants make American society better in the long run is the same for those living in communities with immigrants and those living without immigrants in their communities. Another way to state this is: We assume that the difference in proportions between those who live near immigrants and those who do not is zero.

$$H_0 : p_{incommunity} - p_{notincommunity} = 0.$$

The alternative hypothesis is that there is a difference in attitudes.

$$H_A : p_{incommunity} - p_{notincommunity} \neq 0.$$

This will be conducted as a two tailed test because we do not have a theory in advance about which group will have a higher opinion of immigrants.

- What is the difference between the “in community” and “not in community” sample proportions?
- In a hypothesis test for the difference between two proportions we are assuming that the two proportions are equal. In order to get a standard error under this assumption we need to get a pooled proportion. In this case, that means that we should find the proportion of the survey respondents who indicated yes for both types of communities combined. What is the pooled proportion?

C. Calculate the pooled standard error for this hypothesis test.

$$SE_{pooled} = \sqrt{\hat{p}_{pooled}(1 - \hat{p}_{pooled}) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)},$$

where $\hat{p}_{pooled} = \frac{x_1 + x_2}{n_1 + n_2}$.

- D. Calculate a test statistic (z -score) for the difference we see between the “in community” and “not in community” groups. Recall that the z -score tells you how many standard deviations (or SEs) your sample data is from the expected difference. The expected difference is stated in the null hypothesis. Clearly label your z -score. *Example:* z -score = 3.8 SEs.
- E. Calculate the p -value for this hypothesis test. Because this problem is using a two tailed test, our p -value should tell us the probability of getting sample data as extreme as our data in either direction. This will require you to double the p -value you would usually calculate using `norm.dist()`.
- F. If possible, draw a conclusion from this hypothesis test. Explain what your calculations mean in this particular context.

Problem 3: On your own

According to the same Pew study, about 1 in 4 (24%) Americans say they have a friend or relative who is a recent immigrant, while those numbers are 1 in 3 (33%) for Americans with college degrees. For “younger Americans”, ages 18 to 29, 36% say they have a friend or relative who is a recent immigrant. The number of people in each subgroup was not readily available in the article, but we’ll use the following table as an estimate.

	Americans	Americans with college degrees	Younger Americans (ages 18 to 29)
sample size	3100	1200	600
know a recent immigrant	24%	33%	36%

For both of the following, clearly label the steps of your hypothesis test, including

- Stating the null and alternative hypotheses. Both of the following scenarios should be two-tailed tests.
 - Determining the mean and standard error of the sampling distribution.
 - Calculating the p -value for the sample statistics.
 - Drawing conclusions in both statistical language and common language tied to the original context.
- A. Conduct a hypothesis test for the null hypothesis that the proportion of Americans who know a recent immigrant is the same for Americans with college degrees and Americans in general.
- B. Is there a difference in the chance an American will know a recent immigrant depending on whether that American is a college graduate or a younger American?

Before you submit your lab, ask yourself:

- Are your hypothesis tests easy to follow, with each step clearly labeled?
- Are your conclusions written in statistical terms (e.g. reject the null hypothesis because we have a very low p -value) AND phrased in everyday language related to the context of immigration?