CSC 315, Fall 2018
Lab #6: Hypothesis Tests -- Proportions

You should create a <u>single</u> R script that covers all problems, and your script should have a heading similar to the following:

```
###################################################
##  Garrett Dancik
##  CSC 315, Lab #6
###################################################
```

Your answer to each question should be numbered in a comment in your R script. When your script is complete, create a Notebook and turn a hardcopy of the Notebook in when the assignment is due.

1. In a Gallup poll released October 14, 2016, American adults were asked, "Which do you think should be the bigger priority for the U.S. criminal justice system today?", and were given two choices. Of those surveyed, 498 individuals said "strengthening law and order", while 437 individuals answered "reducing bias against minorities". (Approximately 8% of individuals had "no opinion", but we will ignore these responses in our analysis). Let $\hat{p}$ be the proportion of the population who prioritizes "strengthening law and order" over "reducing bias against minorities".

   a. What is the mean and standard deviation of the distribution of $\hat{p}$ under the null hypothesis that $p = 0.50$.

   b. Graph the distribution of $\hat{p}$ and draw a vertical line at $\hat{p} = 498/935$, the proportion preferring "strengthening law and order".

   c. Calculate the z test statistic and graph its distribution under the null hypothesis under the graph in (a) as was done in class, drawing a vertical line at the z statistic. Find the p-value based on this test statistic.

   d. Use the *prop.test* function to conduct the hypothesis test *without* the continuity correction. Calculate the z test statistic from the *prop.test* object and extract the p-value (Note: these should match the test statistic and p-value from part (b).

   e. State the conclusion regarding whether American adults wish to prioritize "law and order" or "reducing bias", based on this poll, and justify your answer based on the p-value. Use a significance level of 0.05.

f. What would it mean in the context of this problem if a Type I error occurred?

2. A person who claims to be a psychic says that she can correctly predict the outcome of a roll of a standard die (with numbers 1-6) more times than what would be expected by chance. When you roll a die 50 times, she correctly predicts the outcome 12 times. (Note: you may use the *prop.test* function to complete (b) and (c).

   a. State the null and alternative hypothesis corresponding to this claim.

   b. Find the z test statistic

   c. Find the p-value

   d. State the conclusion regarding whether or not the person's ability to predict the outcome of the die is different than what would be expected by chance.

   e. What would it mean in the context of this problem if a Type II error occurred?

3. Five years ago, 36% of adults reported making all or most of their purchases with cash. In a Gallup poll conducted June 22-23, 2016, a survey of 1,024 adults found that 246 (~24%) of those surveyed make all or most of their purchases with cash. Is there evidence that the proportion of adults making all or most of their purchases in cash has changed in the past 5 years? Answer this question by carrying out steps (a) – (d) of the problem above, testing specifically whether the current proportion differs from 36%.

4. Find the p-values associated with the following z test statistics, and state whether you would reject or fail to reject the null hypothesis at $\alpha = 0.05$.

   a. $z = 3.32$ b. $z = -1.3$ c. $z = -2.02$

5. A study was conducted to evaluate whether or not a daily dose of aspirin could reduce the risk of dying from cancer. The study used a randomized trial design, where individuals were randomly assigned to receive aspirin or a placebo. (Note: This randomization allows one to conclude a cause-and-effect relationship between the treatment and the response variable). The results of the study are in the table below. Note that you can use the *prop.test* function for (d) – (f).

Group	Death from Cancer		
	Yes	No	Total
Placebo	347	11188	11535
Aspirin	327	13708	14035

a. State the null and alternative hypothesis (*this is completed for you*)

$$H_0: p_{aspirin} - p_{placebo} = 0$$
$$H_A: p_{aspirin} - p_{placebo} \neq 0$$

where $p_{aspirin}$ is the probability an individual receiving aspirin will die from cancer, and $p_{placebo}$ is the probability an individual receiving the placebo will die from cancer

b. What is the estimate of the common population proportion (i.e., the value of p on slide 18).

c. What is the standard deviation of the difference between the two proportions (i.e., the value of the denominator of Z on slide 18)?

d. What are the sample proportions? (You can use *prop.test* to get this)

e. What is the z test statistic?

f. What is the p-value?

g. State the conclusions regarding whether or not aspirin can prevent the rate of deaths due to cancer, and justify your conclusions based on the p-value.

h. What would it mean in the context of this problem if a Type I error occurred?

i. What would it mean in the context of this problem if a Type II error occurred?