STA 6126 – HW#2 – Due October
Confidence Intervals and Significance Tests

Part 1: Confidence Intervals for Means, Proportions, and Medians

- Download the Movie Revenue/Rotten Tomatoes Score Dataset and print out the first 6 rows.
- Obtain the population mean, median, and standard deviation of revenues
- Obtain the population proportion of Fresh (>= 0.600 proportion of postive reviews among all reviewers)
- Take 10000 random samples of n = 36 films. Obtain 95% Confidence Intervals for the population mean and median revenue, and proportion of Fresh. Give the coverage rates for the 3 sets of Confidence Intervals.
- Give the theoretical and empirical means and standard errors for the sample mean revenues and proportion Fresh.

Part 2: Significance Tests for a Mean and a Proportion

- In a class with 53 students, each student completed 25 "ESP tasks" to predict which of 3 images had been selected as the target image. Each student saw the 3 images, then the monitor was muted and one of the 3 images (which had been selected via a random number generator) was selected. The students then guessed which of the 3 images was the target. Note that there were a total of 53(25) = 1325 choices made, and each choice could be either correct or incorrect. There were 484 correct selections.

    - What is the parameter of interest, μ or π?
    - What is the parameter value under the null hypothesis, assuming no ESP (that is, that all guesses are purely random)?
    - What is the standard error of the estimator under the null hypothesis?
    - What is the point estimate of the parameter from the experiment?
    - Compute the test statistic.
    - Assuming an upper tail test (evidence in favor of ESP), what is the P-value of the test?

- For the Stroop reading experiment, results separate by gender are given below, where Y is the difference in time to read list in different color minus list in black. Conduct the tests for interference separately for females and males. For each test, give the null and alternative hypotheses, test statistic and P-value. Note that for males, you will need to make use of the t-distribution.

Gender	n	Mean	SD
Female	56	2.28	7.99
Male	14	2.41	7.05

Part 3: Comparing 2 Means – Independent Samples

```
Dataset:  kid_calories.csv

Source: K. nan der Horst, A. Ferrage, A. Rytz (2014). "Involving Children
in Meal Preparation," Appetite. Vol. 79, pp.18-24

Description: Experiment conducted to determine effect of children
participating in meal preparation on caloric intake. Two groups
(independent samples). Treatment 1: Children participated.
Trt 2: Did not participate. Data generated to match means and SDs.

Variables:
Treatment
Caloric Intake
```

This experiment randomly assigned children to either participate in meal preparation (Treatment 1) or not participate (Treatment 2). Conduct the independent sample t-test to compare the treatments with respsct to calorie intake (Y) by completing the following parts.

- Give the means, standard deviations, and sample sizes for the two groups.
- Give the point estimate for $\mu_1 - \mu_2$ and its estimated standard error (equal variances assumed)
- Give the null and alternative hypotheses for a 2-sided test
- Give the test statistic, rejection region ($\alpha = 0.05$) and P-value
- Compute a 95% Confidence Interval for $\mu_1 - \mu_2$

Part 4: Comparing 2 Proportions – Independent Samples

A research paper studied a random sample of 50 slasher horror movies. They classified characters based on gender (Female/Male) and Sexual Activity (Yes/No). Further they determined whether the character survives (Yes/No). Consider the two groups: Female/Sex=Yes (Group 1) and Male/Sex=No (Group 2). In group 1, there were $n_1 = 83$ characters with $x_1 = 11$ surviving. In group 2, there were $n_2 = 189$ characters with $x_2 = 28$ surving. Complete the following parts (treating these characters as a random sample from the population of all slasher horror movie characters).

- Obtain the point estimate and 95% CI for $\pi_1 - \pi_2$
- Test $H_0: \pi_1 - \pi_2 = 0$ versus $H_A: \pi_1 - \pi_2 \neq 0$ by giving the Test Statistic and P-value.

Part 5: Comparing 2 Proportions – Paired Samples

A study was conducted to study the effects of framing of decisions in terms of gains or losses. The study was conducted on 150 college students who were given each of the following two decisions (Note that this is an example of dependent samples since each subject was given both decisions).

Decision 1 Choose Between:

A. A sure gain of $240

B. 25% Chance for $1000 gain, 75% Chance for $0 gain

Decision 2 Choose Between:

C. A sure loss of $750
D. 75% Chance to lose $1000, 25% Chance to lose $0

The dataset is given in **riskgamble.csv** and is coded such that a 1 represents the respondent selected the "sure thing" and 2 represents selection of the "risky choice". Run McNemar's test to test whether people tend differ in terms of their propensity to select the risky choice when it is framed as a gain versus a loss. That is, test $H_0: \pi_G - \pi_L = 0$ vs $H_A: \pi_G - \pi_L \neq 0$ where π_G , π_L are the true population proportions of individuals selecting the risky choice when decision is based as a gain and loss, respectively.