Discussion Section #11
Due: To be submitted to CatCourses by 11:59pm.

Instructions:

This lab, we will be doing computational exercises which follow Homework #12 and the concepts of hypothesis testing and t Tests.

Please follow along with your instructor as you go through an examples of each. For this assignment answer the questions in the following section. When specified you should include a figure and/or you R code.

Assignment:

1. **(2 Points) Hypothesis Testing.** A certain type of automobile is known to sustain no visible damage 25% of the time in 10-mph crash tests. A modified bumper design has been proposed in an effort to increase this percentage. Let p denote the proportion of all 10-mph crashes with this new bumper that result in no visible damage. Your goal is to look at collected data and conduct a hypothesis test to decide between:

   - H_0: $p = 0.25$ (no improvement from the new bumper)
   - H_1: $p > 0.25$ (the new bumper is better)

   The file Lab11Prob1.csv contains data from N different crash tests. The data for each test is be a binary output, either there is no-damage ($1 =$ success) or damage ($0 =$ no success). Let X_i be the outcome from each experiment, and let T be the total number of successful (i.e., no damage) crash tests:

   $$T = \sum_{i=1}^{N} X_i.$$

   Your goal will be to evaluate T and make a decision about whether or not to reject H_0 at the significance level $\alpha = 0.05$. Notice that this model follows the first example you worked through in the presentation and T behaves according to a binomial distribution.

   (a) (1 Point) Modify Lab11.R to conduct the appropriate hypothesis test. In addition, ensure your command returns a 95% confidence interval in the success probability.

   (b) (1 Point) The data in `Lab11Prob1.csv` are the outcomes of the different experiments. Based on these results should you conclude that the modified bumper design was successful? That is, should you reject the null hypothesis at the $\alpha = 0.05$ level of significance? And what is the confidence interval that R calculates for the true success probability p?

2. **(3 Points) Hypothesis Testing.** A manufacturer of sprinkler systems used for fire protection in office buildings claims that the true average system-activation temperature is $130°F$. You are given a sample of n activation temperatures (Lab11Prob2.csv). You are told that the standard deviation is known to be $1.5°F$.

You are tasked with determining if the null hypothesis H_0 that the true mean of activation temperatures agrees with the manufacturer's reported value of $130°F$ at $\alpha = 0.01$ level of significance.

Because we are trying to verify a manufacturer value, we want a two-sided hypothesis test. That is, we are comparing:

- H_0: $\mu = \mu_0 = 130$
- H_1: $\mu \neq 130.$

(a) (1 Point) Calculate your sample average and convert it to an appropriate z statistic.

(b) (1 Point) Since we have a two-sided hypothesis test, we want to reject the null hypothesis if

$$z \geq z_{\alpha/2} \text{ or } z \leq -z_{\alpha/2}.$$

What are the critical values of $z_{\alpha/2}$? If you used R, provide your R code for calculating them.

(c) (1 Point) Should we reject the null hypothesis at this level of significance?

3. **(2 Points). One-Sample t-Test.** The Academy of Motion Picture Arts and Sciences has determined that the mean run time of a short film produced by an independent film-maker was normally distributed with mean $\mu_0 = 15$ minutes. They are curious if the 2020 pandemic/quarantine has caused a change in behavior. You received a set of runtimes for N independent short films produced during the last few months (Lab11Problem3.csv).

   Because we do not know the variance of our data set, we will have to conduct a t-test to decide between

   - H_0: $\mu = \mu_0$
   - H_1: $\mu < \mu_0$

   (a) (0.5 Points.) Determine the appropriate R code to conduct your one-sample hypothesis test. (Remember, this is a one-sided hypothesis test!)

   (b) (0.5 Points.) How many degrees of freedom does your t-distribution have? Explain what this should be from your data and verify the result by looking at the output of R's hypothesis test.

   (c) (0.5 Points.) Should you reject your null hypothesis?

   (d) (0.5 Points) Looking at the Box Plot of the results, do you agree with the hypothesis test or not? There is no wrong or right answer here, this is just asking you to take a look at the outcome of your test and to see if you think it makes sense. Please include a copy of the Box Plot from R in your write-up.

4. **(3 Points). Two-Sample t-Test.** In this example, you will consider three categories and conduct two-sample t-tests between each pair.

   The duration of pregnancy was measured for 1669 women who gave birth in a maternity hospital in Newcastle-upon-Tyne, England, in 1954. The durations are measured in complete weeks from the beginning of the last menstrual period until delivery. The pregnancies are divided into three categories according to how admission to the hospital was processed. Either a patient was admitted for standard (non-emergency) delivery, emergency delivery, or an extended stay for women who were admitted to the hospital before going into labor due to lack of housing.

The data (mean, $\overline{x}$ and sample variance, s^2) for the three groups were recorded as follows:

1. Standard: 775 patients with $\overline{x} = 39.08$ and $s^2 = 7.77$.
2. Emergency: 261 patients with $\overline{x} = 37.59$ and $s^2 = 25.33$.
3. Extended Stay: 633 patients with $\overline{x} = 39.60$ and $s^2 = 4.95$.

From the widely varying sample variances, researchers conclude the samples have un-equal variance. They were interested in if the average duration in pregnancy differs between the three groups.

For each pair of groups test the null hypothesis of equal means according to a two-sample t test at the $\alpha = 0.05$ level of significance. You will receive 1 point for each of the 3 comparisons. For full credit you must include:

- The specific R command you used.
- Whether or not you reject the null hypothesis.
- A sentence (or more) about whether or not you agree with the hypothesis test by looking at distributions on box plots (or other plots). Please include any plot you use in your decision making.