Name:
GitHub Username:
Purdue Username:
Instructor:
Section:

Problem 1

1.

Given the probability density function:

$$f_X(x) = \begin{cases} \frac{3}{4}(1 - x^2) \ when \ x \in [-1,1] \\ 0 \qquad otherwise \end{cases}$$

If you happen to plot this distribution, then by looking at the plot, determine by inspection, what is the mean of the distribution.

Mean of the distribution:

From this distribution we grab 100 i.i.d samples $X_1, X_2,\dots X_{100}$.

The sample mean random variable is defined as

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

What is the shape of the probability density function of the sample mean?

What is the mean of the probability density function of the sample mean?

Explain how you know the shape and mean.

Part B

Let X be a normally distributed random variable with mean = 0 and variance = 1. What is the distribution of the random variable Z = 3X+1 (Give your answer with the shape, mean and variance).

What is the shape of the distribution?

What is the mean?

What is the variance?

Problem 2

You are given data in *city_vehicle_survey.txt* representing the average age of vehicles across various counties. The transportation department claims the average age of vehicles is 5 years. You are tasked with testing this claim.

1. Formulate null and alternative hypotheses for a statistical test that seeks to challenge this belief. What are the null and alternative hypotheses?

   1. Null Hypothesis:

   2. Alternative Hypothesis:

   3. What type of test should be used and why?

2. Carry out this statistical test using the *city_vehicle_survey.txt* sample. Report the sample size, the sample mean, the standard error, the standard score (z or t, depending on what was used), and the p-value.

   ******ROUND ALL DECIMAL VALUES TO 4 DECIMAL PLACES****

Sample size	
Sample mean	
Standard error	
Standard score	
p – value (if less than 0.01 use scientific notation)	

   Are the results statistically significant at a level of 0.05?

   Yes   **No**

   What (if anything) can we conclude about the hypothesis at the confidence level of 0.05?

Are the results statistically significant at a level of 0.10?

Yes   **No**

What (if anything) can we conclude about the hypothesis at the confidence level of 0.10?

3. What is the largest standard error for which the test will be significant at a level of 0.05? What is the corresponding minimum sample size? (You may assume that the population variance and mean does not change.)

   ******ROUND ALL DECIMAL VALUES TO 4 DECIMAL PLACES****

Largest standard error	
Corresponding minimum sample size	

4. Suppose the transportation department believes the mean vehicle age is the same in counties with and without emission control programs. Two datasets, *vehicle_data_1.txt* (with emission programs) and *vehicle_data_2.txt* (without emission programs), are used to test this assumption.

   1. Null Hypothesis:

   2. Alternative Hypothesis:

   3. What type of test should be used and why?

5. Carry out this statistical test using the *vehicle_data_1.txt* population and *vehicle_data_2.txt* population samples. Report the sample sizes, the sample means, the standard error, the z-score, and the p-value. Are the results significant at levels 0.05 or 0.10? What (if anything) can we conclude about the hypothesis at the two different confidence levels?
   ******ROUND ALL DECIMAL VALUES TO 4 DECIMAL PLACES**** ****

Sample size of *vehicle_data_1* (Emission)	
Sample size of *vehicle_data_2* (Without Emission)	
Sample mean of *vehicle_data_1* (Emission)	
Sample mean of *vehicle_data_2* (Without Emission)	
Standard error	
Standard score	
p – value (if less than 0.01 use scientific notation)	

1. Are the results statistically significant at a level of 0.05?

   Yes   **No**

2. Are the results statistically significant at a level of 0.10?

   Yes   **No**

3. What (if anything) can we conclude (i.e., what is the interpretation of the result)?

Problem 3

1. Use the sample to construct a 90% confidence interval for the average sodium of snacks. Report whether you will use a z-test or t-test and report the sample mean, the standard error, the standard statistic (t or z value), and the interval. (Think, which distribution should you use here if very few data points are available?)

******ROUND ALL DECIMAL VALUES TO 4 DECIMAL PLACES****

Sample mean	
Standard error	
Standard score (t or z value)	
90% confidence interval	

2. Repeat Q1 for a 95% confidence interval.
******ROUND ALL DECIMAL VALUES TO 4 DECIMAL PLACES****

Standard error	
Standard score (t or z value)	
95% confidence interval	

Is your interval wider or narrower compared to using the 90% confidence interval in Q1?
Wider   **Narrower**

3. Repeat Q2 if you are told that the population standard deviation is 5.
   Will you use a t-test or z-test (Hint: Think which distribution should you use here now that you have the true population standard deviation)? Justify your answer.

******ROUND ALL DECIMAL VALUES TO 4 DECIMAL PLACES****

Standard error	
Standard score (t or z value)	
95% confidence interval	

Is your interval wider or narrower than the interval computed in Q2?
Wider   **Narrower**