

Name:

GitHub Username:

Purdue Username:

Instructor:

Section:

Problem 1

Part A

Given the probability density function:

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

Use Python to plot the probability density function from -1 to 1. When 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

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

We know that the variance of each X_i is $\frac{3}{5}$.

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 $Y = -X$ (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

A student exploring making a car maintenance subscription service an "uber for oil changes" is pretty confident that the average age of used cars on campus (the vehicle_survey_1.txt population) is 8 years old.

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

Null Hypothesis:

Alternative Hypothesis:

What type of test should be used and why?

2. Carry out this statistical test using the vehicle_survey_1 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) | |

(For the following fill the box next to your selected answer)

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?

(For the following fill the box next to your selected answer)

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 student is also convinced that the mean age is different between vehicles whose owners regularly change their own oil (the vehicle_survey_2 population) and those who do not (the vehicle_survey_3 population). Formulate null and alternative hypotheses that seek to validate this belief. What are the null and alternative hypotheses, and what type of test can be used?

Null Hypothesis:

Alternative Hypothesis:

What type of test should be used and why?

5. Carry out this statistical test using the `vehicle_survey_2` and `vehicle_survey_3` 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_survey_2 (non oil changers) | |
| Sample size of vehicle_survey_3 (oil changers) | |
| Sample mean of vehicle_survey_2 (non oil changers) | |
| Sample mean of vehicle_survey_3 (non oil changers) | |
| Standard error | |
| Standard score | |
| p – value (if less than 0.01 use scientific notation) | |

Problem 3

1. Use the sample to construct a 90% confidence interval for the weight of the robots on average. 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 datapoints are available?)

Will you use a t-test or z-test (Hint: Think which distribution should you use here if very few data points are available)? Justify your answer.

| |
|--|
| |
|--|

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

| | |
|-------------------------------|--|
| Sample mean | |
| Standard error | |
| Standard score (t or z score) | |
| 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 score) | |
| 95% confidence interval | |

(For the following, fill the box next to your selected answer)

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 10.

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 score) | |
| 95% confidence interval | |

Is your interval wider or narrower than the interval computed in Q2? (Fill Box to select your answer)

Wider

Narrower