Name: Alex Yan

Instructor: Qiang

GitHub Username: PebbleBro

Purdue Username: ajyan

Problem 1

1. Suppose the instructor of the course is convinced that the mean engagement of students who become knowledgeable in the material (i.e., the engagement\_1 population) is 0.75.
   1. *[5 points]* Formulate null and alternative hypotheses for a statistical test that seeks to challenge this belief. What are the null and alternative hypotheses?

**ANSWER:**

H0: The mean engagement of students who became knowledgeable in the material is 0.75

Ha: The mean engagement of students who became knowledgeable in the material is not 0.75

* 1. *[5 points]* What type of test should be used and why?

**ANSWER:**

We can use the z-test because the sample size is greater than 30 (Central Limit Theorem)

1. Carry out the statistical test defined in (1b) using the *`engagement\_1`* sample.
   1. *[1 point]* What is the sample size?

**ANSWER:** 937

* 1. *[1 point]* What is the sample mean?

**ANSWER:** 0.743

* 1. *[2 points]* What is the standard error?

**ANSWER:** 0.127

* 1. *[2 points]* What is the standard score?

**ANSWER:** -0.055

* 1. *[2 points]* What is the p-value?

**ANSWER:** 0.956

* 1. *[2 points]* Are the results statistically significant at a level of 0.05? How about 0.10? What (if anything) can we conclude (i.e., what is the interpretation of the result)?

**ANSWER:** The results are not statistically significant at both levels. Hence, we fail to reject the null hypothesis.

1. *[10 points]* 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.)

**ANSWER:** Standard error = 0.0134 Sample size = 84275

1. Suppose the instructor is also convinced that the mean engagement is different between students who become knowledgeable (the engagement\_1 population) and those who do not (the engagement\_0 population).
   1. *[5 points]* Formulate null and alternative hypotheses that seek to validate this belief. What are the null and alternative hypotheses?

**ANSWER:**

H0: The mean engagement is the same between students who became knowledgeable and those who do not.

Ha: The mean engagement is different between students who became knowledgeable and those who do not.

* 1. *[5 points]* What type of test should be used and why?

**ANSWER:**

We can use the z-test because the sample size is greater than 30 (Central Limit Theorem)

1. Carry out the statistical test defined in (4b) using the *`engagement\_0`* and *`engagement\_1`* samples.
   1. *[1 point]* What are the sample sizes?

**ANSWER:**

Sample size of engagement 0 = 1977

Sample size of engagement 1 = 937

* 1. *[1 point]* What are the sample means?

**ANSWER:**

Sample mean of engagement 0: 0.640

Sample mean of engagement 1: 0.743

* 1. *[2 points]* What is the standard error?

**ANSWER:**

Standard deviation of engagement 0: 0.254

Standard deviation of engagement 1: 0.127

* 1. *[2 points]* What is the standard score?

**ANSWER:**

Z-score = -3.652584616837238

* 1. *[2 points]* What is the p-value?

**ANSWER:**

P-value = 0.00025961400658740616

* 1. *[2 points]* Are the results statistically significant at a level of 0.05? How about 0.10? What (if anything) can we conclude (i.e., what is the interpretation of the result)?

**ANSWER:**

Yes, the results are statistically significant at both levels of 0.05 and 0.10. Thus, we reject the null hypothesis.

Problem 2

1. Use the sample to construct a 90% confidence interval for the number of points by which the team wins on average.
   1. *[3 points]* 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.

**ANSWER:**

A t-test will be used here because there are less than 30 data points available and the population standard deviation is unknown.

* 1. *[3 points]* What is the sample mean?

**ANSWER:**

Sample mean = 7.36

* 1. *[3 points]* What is the standard error?

**ANSWER:**

Standard error = 16.05

* 1. *[3 points]* What is the standard statistic (t or z value)?

**ANSWER:**

t-value = 1.81

* 1. *[3 points]* What is the 90% confidence interval?

**ANSWER:**

90% Confidence interval = (-21.73,36.46)

1. Repeat Q1 for a 95% confidence interval.
   1. *[2 points]* What is the standard statistic (t or z value)?

**ANSWER:**

t-value = 2.23

* 1. *[2 points]* What is the 95% confidence interval?

**ANSWER:**

95% confidence interval = (--28.40,43.13)

* 1. *[1 point]* Is your interval wider or narrower compared to using the 90% confidence interval in Q1?

**ANSWER:**

The 95% confidence interval is wider than the 90% confidence interval.

1. Repeat Q2 if you are told that the population standard deviation is 15.836.
   1. *[5 points]* 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.

**ANSWER:**

Z-test because we know the population standard deviation.

* 1. *[3 points]* What is the standard error?

**ANSWER:**

Standard error = 4.77

* 1. *[3 points]* What is the standard statistic (t or z value)?

**ANSWER:**

Z-score = 1.96

*[3 points]* What is the 95% confidence interval?

**ANSWER:**

95% confidence interval = (-1.00, 16.72)

* 1. *[6 points]* Is your interval wider or narrower than the interval computed in Q2?

**ANSWER:**

This interval is narrower than the interval computed in Q2.

1. *[10 points]* Assume you no longer know the population standard deviation. With what level of confidence can we say that the team is expected to win on average? (Hint: What level of confidence would you get a confidence interval with the lower endpoint being 0?)

**ANSWER:**

With a 34% level of confidence we can say the team is expected to win on average.