Name: Connor Damato

Instructor: Rajeev Sahay

GitHub Username: Connor-Damato

Purdue Username: damato0

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:**

**Null hypothesis: The true mean engagement of students is 0.75**

**Alternative hypothesis: The true mean engagement of students is not 0.75**

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

**ANSWER: A Z-Test should be used, since we have more than 30 samples and we are comparing a true mean to a sample mean.**

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

**ANSWER:** 931

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

**ANSWER:** 0.7427

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

**ANSWER:** 0.004170

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

**ANSWER:** -1.7567

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

**ANSWER:** 0.0790

* 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: No, they are not statistically significant at a level of 0.05. They are significant at a level of 0.10. With a significance level of 0.05, we are unable to reject the null hypothesis that the true mean engagement of students is 0.75**

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:**

SE- 0.003738

N = 1161

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:**

**Null: The true mean of engagement 1 is the same as engagement 0**

**Alternative: The true mean of engagement 1 is different than engagement 0**

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

**ANSWER: A 2 sample Z test should be used**

1. Carry out the statistical test defined in (4b) using the *`engagement\_1`* and *`engagement\_2`* samples.
   1. *[1 point]* What are the sample sizes?

**ANSWER:**

Engagement\_0- 1970

Engagement 1- 931

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

**ANSWER:**

Engagement\_0- 0.6396

Engagement 1- 0.7427

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

**ANSWER:**

SE -0.007090

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

**ANSWER:** -14.531463955605583

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

**ANSWER: 7.656643244263393e-48**

* 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. We can reject the null hypothesis and state that the true means of the two samples are different.

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: t-test**

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

**ANSWER:** 6.6923076923076925

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

**ANSWER:** 4.359464527492962

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

**ANSWER:** 1.782287555649159

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

**ANSWER:** (-1.0775116843369554, 14.462127068952341)

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

**ANSWER:** 2.1788128296634177

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

**ANSWER:** (-2.806149550656543, 16.19076493527193)

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

**ANSWER:** Wider

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, we know the true population standard deviation so we can use z despite not having a large enough n.**

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

**ANSWER:** 4.392116153711364

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

**ANSWER:** 1.959963984540054

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

**ANSWER:** (-1.9160817848831693, 15.300697169498555)

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

**ANSWER:** Narrower

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:** 0.8493106444354639