Name: Calvin Madsen
Instructor: Qiang Qiu
GitHub Username: MadSons
Purdue Username: cfmadsen

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.

   a. *[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 (H_0): The instructor is correct, $\mu = 0.75$
      Alternative hypothesis (H_1): The instructor is incorrect, $\mu \neq 0.75$

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

      ANSWER:
      Z-Score should be used as there are more than 29 data points (there are 931)

2. Carry out the statistical test defined in (1b) using the *`engagement_1`* sample.

   a. *[1 point]* What is the sample size?

      ANSWER: 931

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

      ANSWER: 0.7426735936745436

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

      ANSWER: 0.004172740741992158

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

      ANSWER: -1.755777983455224

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

      ANSWER: 0.07912632510370125

f. *[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: $p = 0.079$. At a level of 0.05, $p > 0.05$ and for 0.10, $p < 0.1$. Not significant for a level of 0.05 (cannot reject H_0), but is significant for a level of 0.1 (can reject H_0) .

3. *[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:
   Largest Standard Error for significance of 0.05: 0.003738031098145784
   Corresponding minimum sample size: 1161

4. 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).

   a. *[5 points]* Formulate null and alternative hypotheses that seek to validate this belief. What are the null and alternative hypotheses?

   ANSWER:
   Null hypothesis (H_0): The instructor is incorrect, μ _1 (knowledgeable) = μ _0 (not knowledgeable)

   Alternative hypothesis (H_1): The instructor is correct, μ_1 (knowledgeable) ≠ μ_0 (not knowledgeable)

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

   ANSWER: two-sample-z-test because we are testing two different population means and we have more than 29 data points.

5. Carry out the statistical test defined in (4b) using the *`engagement_1`* and *`engagement_2`* samples.

   a. *[1 point]* What are the sample sizes?

   ANSWER:
   Sample Size engagement_0: 1970
   Sample Size engagement_1: 931

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

ANSWER:
Sample Mean engagement_0: 0.6396449134177665
Sample Mean engagement_1: 0.7426735936745436

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

ANSWER: 0.007090041345561141

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

ANSWER:
Standard Score: -14.531463955605863

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

ANSWER 7.656643244232126e-48

f. *[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: Results are significant at a level of 0.05 and 0.1 as p is far less than both numbers and is practically zero. This indicates that we can reject the null hypothesis of the two means being equal, so the instructor was correct, and the population means between the two data sets cannot be equal. So,
μ_1 (knowledgeable) $\neq$ μ_0 (not knowledgeable).

Problem 2

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

    a. *[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 because there are fewer than 30 data points

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

       ANSWER: 6.6923076923076925

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

       ANSWER: 4.359464527492962

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

       ANSWER: 1.782287555649159

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

   ANSWER: (-1.0775116843369554, 14.462127068952341)

2. Repeat Q1 for a 95% confidence interval.

   a. *[2 points]* What is the standard statistic (t or z value)?

      ANSWER: 2.1788128296634177

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

      ANSWER: (-2.806149550656543,16.19076493527193)

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

      ANSWER: Wider due to it needing to incorporate 95% confidence instead of 90%.

3. Repeat Q2 if you are told that the population standard deviation is 15.836.

   a. *[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 have true population standard deviation.

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

      ANSWER: 4.392116153711364

   c. *[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)

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

      ANSWER: Narrower, z test is more confident as we have actual population deviation.

4. *[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:

C = 0.8465080318837064