Name: Joao Taff-Freire
Instructor: Mahsa Ghasemi
GitHub Username: Frucks
Purdue Username: jtafffre

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: The null hypothesis is H0: mu = 0.75. The alternative hypothesis is H1: mu != 0.75

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

    ANSWER: We should use a Z-score test, as we have at least 30 data points.

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

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

    ANSWER: 937

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

    ANSWER: 0.7430304110448239

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

    ANSWER: 0.004153027288269652

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

    ANSWER: -1.6781948375012814

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

    ANSWER: 0.09330906925243751

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: At a level of 0.05, we failed to reject the null hypothesis for the 95% confidence level and the p-value is statistically significant.

At a level of 0.10, we can reject the null hypothesis for the 90% confidence level and the p-value is not statistically significant.

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: At a level of 0.05, the largest standard error for which the test will be significant is: 0.004237209220916103
The minimum sample size is: 900

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: The null hypothesis is that the mean engagement is 0.75. The alternative hypothesis is that the mean engagement is different between the two.

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

ANSWER: The z-test should be used, as the sample size is bigger than 30.

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: 1977 and 937

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

ANSWER: 0.6399545077035914 and 0.7430304110448239

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

ANSWER: 0.005715989588773277

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

ANSWER: -19.252220562568542

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

ANSWER: 1.352401028607762e-82

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: We can reject the null hypothesis for the 95% confidence level, and the p-value is not statistically significant.
We can reject the null hypothesis for the 90% confidence level, and the p-value is not statistically significant.

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, as there are fewer than 30 data points.

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

    ANSWER: 7.363636363636363

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

    ANSWER: 4.840040856450614

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

    ANSWER: 1.4308981556180451

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

    ANSWER: (0.4380308290251964, 14.28924189824753)

2. Repeat Q1 for a 95% confidence interval.

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

ANSWER: 1.4407478926091581

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

ANSWER: (0.39035769956291677, 14.33691502770981)

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

ANSWER: My interval is wider compared to the 90% interval.

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: Since we know the standard deviation, I will use a z-score.

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

ANSWER: 4.774733652733465

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

ANSWER: 1.6448536269514722

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

ANSWER: (-0.49010160278952775, 15.217374330062254)

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

ANSWER: My interval is wider than the interval computed in Q2.

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: We can say that the team is expected to win on average with a confidence level of 84%