Name: Roshan Sundar
Instructor: Mahsa Ghasemi
GitHub Username: roshsundar
Purdue Username: rmsundar

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: H_0: $\mu = 0.75$, H_1: $\mu \neq 0.75$

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

    ANSWER: The z-test (hypothesis test) should be used since a claim is trying to be proved/disproved.

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

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

    ANSWER: 1970

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

    ANSWER: 0.6396

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

    ANSWER: 0.0057

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

    ANSWER: 19.2521

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

    ANSWER: 0.0000

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: since the p-value is less than 0.05 and 0.10, the results are significant in both cases. Therefore, the hypothesis can be rejected.

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: SE = 0.0563, n = 21

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: H_0: $\mu_0 = \mu_1$, H_1: $\mu_0 \neq \mu_1$

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

ANSWER: The two-sample z-test should be used – since we are trying to prove/disprove claim and compare two population means as well.

5. Carry out the statistical test defined in (4b) using the *`engagement_0`* and *`engagement_1`* samples.

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

ANSWER: n_0: 1970, n_1: 931

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

ANSWER: μ_0: 0.6396, μ_1: 0.7427

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

ANSWER: 0.0071

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

ANSWER: 14.5315

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

ANSWER: 0.0000

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: since the p-value is less than 0.05 and 0.10, the results are significant in both cases. Therefore, the hypothesis can be rejected.

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: A t-test will be used since the number of data points is small (less than 30)

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

    ANSWER: 6.6923

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

    ANSWER: 4.3595

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

    ANSWER: 1.7823

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

    ANSWER: (-1.0775, 14.4621)

2. Repeat Q1 for a 95% confidence interval.

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

    ANSWER: 2.1788

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

    ANSWER: (-2.8061, 16.1908)

    c. *[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.

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: A z-test will be used since the standard deviation is already known.

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

    ANSWER: 4.3921

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

    ANSWER: 1.96

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

    ANSWER: (-1.9161, 15.3007)

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

    ANSWER: This interval is narrower than the interval 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: 84.9311%