

Name: Shao Ning Huang
Instructor: Mahsa Ghasemi
GitHub Username: ShaoNingHuang
Purdue Username: huan1645

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: We should use z test because the sample size is large enough

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: no, it is not significant at level of 0.05, while it is significant at 0.1.
we cannot reject the result at level 0.05, but we can reject the test at 0.1.

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: maxstderror :0.00423720922091610, min samplesize: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: $H_0: \mu_0 = \mu_1$
 $H_1: \mu_0 \neq \mu_1$

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

ANSWER: z-test should be used because the sample size is large enough

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: samplesize0: 1977
samplesize1: 937

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

ANSWER: samplemean0:0.6399545077035914
samplemean1:0.7430304110448239

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

ANSWER: 0.007065420910043284

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

ANSWER: -14.588784540028351

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

ANSWER: 3.3104307168195455e-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: yes, it's significant at both level. meaning that we can reject at both level.

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 should be used because the sample size is less than 30

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

ANSWER: 7.363636363636363

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

ANSWER: 5.0762776757504415

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

ANSWER: tscore=1.8124611228107335

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

ANSWER: [-1.8369195722533416, 16.56419229952607]

2. Repeat Q1 for a 95% confidence interval.

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

ANSWER: tvalue: 2.2281388519649385

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

ANSWER: [-3.9470151490654715, 18.674287876338198]

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

ANSWER: wider

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 should be used because we know the population standard deviation

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

ANSWER: 4.774733652733465

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

ANSWER: 1.959963984540054

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

ANSWER: [-1.9946696314926058, 16.721942358765332]

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

ANSWER: narrower

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: 91%