Terrence Randall

Homework #5

Problem 1:

<All code used to solve these questions is located in problem1.py>

1.
   H_0 = The mean engagement of students who become knowledgeable in the material is 0.75.
   H_1 = The mean engagement of students who become knowledgeable in the material is NOT 0.75.
   To test this, we can use the Z-test.
2.
   Using the "eng1.txt" data set:
   (sample size) n = 937
   (sample mean) μ = 0.743
   (standard error) SE = 0.127 / √937 = 0.004
   (standard score) z-score = -1.75
   p-value = 0.0933

   The result is statistically for alpha=0.1 (since the p-value is less than 0.1). But it is not statistically significant for alpha=0.05 or 0.01.

   For alpha=0.01 and 0.05 we would have failed to reject H_0.
   But if we had chosen alpha=0.1, we would have rejected H_0 and accepted H_1.
3.
   The largest SE for which the test would be significant at a level of 0.05 is 0.00355.

   The corresponding sample size that would achieve this is n = 1279.
4.
   H_0 = The mean engagement of students who become knowledgeable in the material and those who don't IS THE SAME.
   H_1 = The mean engagement of students who become knowledgeable in the material and those who don't IS DIFFERENT.

   To test this, we can use the two-sample z-test.
5.
   Under the Null Hypothesis the mean and standard deviation is:
   μ = 0
   SD = sqrt([SD_0^2 / n_0] + [SD_1^2 / n_1]) = sqrt([0.0645 / 1977] + [0.0161 / 937]) = 0.007

   N_0 = 1977 ; N_1 = 937
   μ_0 = 0.6399 ; μ_1 = 0.743
   SE_0 = 0.006 ; SE_1 = 0.004

z-score = (μ_0 - μ_1) / SD = (0.6399 – 0.743) / 0.007 = -14.728
p-value ≈ 0 (very close to zero)

From almost any alpha level, we can reject the null hypothesis that both averages are statistically equal and accept the alternative hypothesis.

Problem 2:

<All code used to solve these questions is located in problem2.py>

1.

Using the t-test is more necessary in this example since we're only given 11 data points. (And assuming the population mean is zero [μ=0])

Sample mean = 7.36
SE = s / sqrt(n) = 37.646 / sqrt(11) = 11.351
t-score = 0.648
(With a desired confidence interval of 95%, the t_score_c corresponding this is 2.228)
95% confidence interval = (sample mean – t_score_c *SE) to (sample mean + t_score_c *SE)
= (7.36 – 2.228*11.351) to (7.36 + 2.228*11.351)
= -17.93 to 32.65

2.

Using the same values as the question above, except with a new confidence interval of 90%.
The new t_score_c2 is 1.812

90% confidence interval = (sample mean – t_score_c2 *SE) to (sample mean + t_score_c2 *SE)
= (7.36 – 1.812 *11.351) to (7.36 + 1.812 *11.351)
= -13.208 to 27.92

3.

If given a value for the standard deviation, we can now use the normal distribution to test against. So, we can now use the z-test.
(Still assuming that the population mean is zero [μ=0])

SE = 9.720
z_score = 0.758
95% confidence interval = -11.688 to 26.415

The interval that we get from this test is much tighter than from the first test in question one. But after giving ourselves a wider interval in question 2 it got tighter, however still not as tight as the interval that we achieved here by knowing the standard deviation.

4.

Terrence Randall

Assuming that we again don’t know the standard deviation, to figure out the level of confidence that we can say that we expect the team to win we can continuously try lower confidence levels until the lower bound is zero.

<this process can be seen in problem2 function Q4>

Confidence level that we expect the team to always win = 45.99%

The interval for this confidence level is from 0.162 to 14.566