Practice Assignment: T-Distribution

Golf is a club-and-ball sport in which players use various clubs to hit balls into a series of holes on a golf course in as few strokes as possible. Each player drives a golf ball off a wooden tee at the start and eventually putts the ball on a soft grass surface into a hole. In golf, you are at a considerable advantage if you can drive the golf ball far. As professional golfers gain more experience and technology continues to improve, it makes sense to assume that professional golfers are driving the ball farther year after year. To test this theory, researchers set out to see if professional golfers are driving the golf ball noticeably further than they were three years prior. Data were collected from 111 male professional golf players.[1] Their average driving distances off the tee (in yards) in 2015 and in 2018 were recorded, and the changes in their average driving distances from 2015 to 2018 were calculated (using order of subtraction 2018 – 2015). Call the change in average driving distance from 2015 to 2018 each player's "improvement score."

1) Go to the *DCMP Explore Quantitative Data* tool at https://lumen-learning.shinyapps.io/eda_quantitative/.
   - Open spreadsheet DCMP_STAT_12B_PGA_Data.
   - Select "Your Own" under "Enter Data."
   - Enter a descriptive name for the variable (e.g., Improvement Score).
   - Copy and paste the 111 Improvement Scores and enter the variable name "Improvement" into the tool.

   Part A: Examine the histogram, boxplot, dotplot, and summary statistics of these data. Write a few sentences describing the features of the data distribution. Address the shape, center, spread, and outliers.

   Part B: What is the sample mean improvement score? What is the appropriate symbol for this value?

   Part C: What is the sample standard deviation? What is the appropriate symbol for this value?

[Continued on the next page.]

[1] *PGA tour data*. (2019). Kaggle. Retrieved from https://www.kaggle.com/jmpark746/pga-tour-data-2010-2018

Part D: Which of the following is a correct interpretation of the standard deviation from Part C in the context of the problem?

a) The typical difference between an individual improvement score for a male professional golfer in this sample and the mean improvement score of 3.72 yards is 4.85 yards.

b) About 50% of individual improvement scores for male professional golfers in this sample are within 4.85 yards of the mean improvement score of 3.72 yards.

c) Every male professional golfer in this sample will have an individual improvement score that is 4.85 yards away from the mean improvement score of 3.72 yards.

d) The average improvement score for male professional golfers in this sample is 4.85 yards.

2) Now, let's estimate the sampling variability of the sample mean improvement score across different samples of 111 male professional golfers.

Part A: Use the summary statistics from the sample data to calculate the standard error of the sample mean.

Part B: Which of the following is a correct interpretation of the standard error of the sample mean from Part A?

a) Every sample of 111 male professional golfers will have a mean improvement score that is 0.460 yards away from the population mean improvement score.

b) The average sample mean improvement score across all samples of 111 male professional golfers is 0.460 yards.

c) The typical difference between a mean improvement score in a sample of 111 male professional golfers and the population mean improvement score is 0.460 yards.

d) About 50% of samples of 111 male professional golfers will yield a sample mean improvement score that is 0.460 yards away from the population mean improvement score.

Part C: Explain the key differences in how the interpretation of the standard error in Part B differs from the interpretation of the sample standard deviation in Question 1, Part D.

3) How many standard errors away from 0 is the observed sample mean?

4) Are the conditions met to use a t Distribution to approximate the distribution of the t-statistic? Justify your answer.

5) What degrees of freedom should you use when calculating probabilities from a t Distribution for these data?

6) Assuming there was no true change in male professional golfers' average driving distances from 2015 to 2018 (i.e., the true mean improvement score was equal to 0), use the appropriate t Distribution to calculate the probability of observing the sample mean from Question 1, Part B or something larger.

   What does this result suggest about the theory that professional golfers are driving the ball farther year after year? Explain.

7) If we had collected a random sample of $n = 35$ male professional golfers rather than 111 and observed the same sample mean and sample standard deviation, specify how each of the following quantities would change.

   Part A: The sample standard deviation when $n = 35$ would be _____.

   a) less than the sample standard deviation when $n = 111$
   b) equal to the sample standard deviation when $n = 111$
   c) greater than the sample standard deviation when $n = 111$
   d) We do not have enough information to determine how the sample standard deviation would change.

   Part B: The standard error of the sample mean when $n = 35$ would be _____.

   a) less than the standard error of the sample mean when $n = 111$
   b) equal to the standard error of the sample mean when $n = 111$
   c) greater than the standard error of the sample mean when $n = 111$
   d) We do not have enough information to determine how the standard error of the sample mean would change.

[Continued on the next page.]

Part C: The t-statistic when $n = 35$ would be _____.

a) less than the t-statistic when $n = 111$
b) equal to the t-statistic when $n = 111$
c) greater than the t-statistic when $n = 111$
d) We do not have enough information to determine how the t-statistic would change.

Part D: The probability calculated in Question 6 when $n = 35$ would be _____.

a) less than the probability calculated in Question 6 when $n = 111$
b) equal to the probability calculated in Question 6 when $n = 111$
c) greater than the probability calculated in Question 6 when $n = 111$
d) We do not have enough information to determine how the probability calculated in Question 6 would change.

