STAT 614 - HW 3

Due: Thursday, October 1, 2020 in Blackboard (go to the Homework folder under the Homework/Classwork content area) by 11:59pm.

Instructions: Please type your solutions to these **FIVE** problems and upload the document as a pdf file in Blackboard. There is only one file to submit for this assignment.

Notes: This homework continues our discussions of sampling distributions and statistical inferences, especially using the t-procedures. You will need some concepts that will be discussed in class next week (hence the two-week due date). This is also the last homework before our first exam!!!

1. Triceps skinfold thickness is an upper arm measurement that has been used as a proxy measure of body fat. The table below gives the mean and standard deviations of tricep skinfold thickness (in cm) for two populations of adult males, those with chronic airflow limitation (such as COPD, a type of obstructive lung disease) and those without any airflow limitation. A study comparing tricep skinfold thickness is being planned in these populations using the respective sample sizes (*n*), also given in the last column of the table.

Population	μ	σ	n
Chronic airflow limitation	0.92	0.4	32
No airflow limitation	1.35	0.5	40

   a. Consider a random sample $y_1, y_2, y_3, \ldots, y_n$ from the chronic airflow limitation population with mean μ and standard deviation σ as given in the table. What is the standard deviation of the sample mean, $\bar{y}$? (Note, this is often called the "standard error" of the mean, especially when an estimate for σ is used.)

   b. Assume the Central Limit Theorem is applicable. What does it suggest about potential values of the sample mean, $\bar{y}$, the researchers can expect in their study?

2. A human resources manager for a large company takes a random sample of 50 employees from the company database. She calculates the mean time that they have been employed. She records this value and then repeats the process: She takes another random sample of 50 names and calculates the mean employment time. After she has done this 1000 times, she makes a histogram of the mean employment times. Is this histogram a display of the population distribution, the distribution of a sample, or the sampling distribution of mean?

3. In most software, the default p-values are computed based on a 2-sided alternative hypothesis (H$_A$: $\mu \neq \mu_0$). However, we may want to use a 1-sided alternative in some problems. Hence, we need to be able to compute the correct 1-sided p-values from the reported 2-sided version. Sketch a graph for each of the following to demonstrate that each is the correct procedure. (You can take a photo of your sketch to include in your HW.)

   (a) For H$_A$: $\mu > \mu_0$:

      i. if test statistic $t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$ is positive, (1-sided p-value) = 0.5*(2-sided p-value).

ii. if test statistic $t = \frac{\overline{x} - \mu_0}{s/\sqrt{n}}$ is negative, (1-sided p-value) = 1 – 0.5*(2-sided p-value).

(b) For H_A: $\mu < \mu_0$:

i. if test statistic $t = \frac{\overline{x} - \mu_0}{s/\sqrt{n}}$ is positive, (1-sided p-value) = 1 – 0.5*(2-sided p-value).

ii. if test statistic $t = \frac{\overline{x} - \mu_0}{s/\sqrt{n}}$ is negative, (1-sided p-value) = 0.5*(2-sided p-value).

4. Suppose the following statement is made in the conclusions section of a paper: "A comparison of breathing capacities of individuals in households with low nitrogen dioxide levels and individuals in households with high nitrogen dioxide levels indicated that there is **no difference** in means (two-sided p-value = 0.24)."

   a. Give a reasonable null and alternative hypothesis that was being tested in this scenario. Carefully define the population parameters of interest being tested.
   b. Why is this statement an inaccurate summary of the hypothesis test?
   c. Re-write the statement so that it is properly summarizing the results of the hypothesis test.

5. **Use the lead.csv data set from HW 1.** Revisit HW 1 for a description of this data set and study. For this problem you will compare the Wechsler full-scale IQ scores (the variable IQF) between the different lead exposure groups, denoted by the GROUP variable.

   a. Compute the mean, standard deviation, standard error, and 95% confidence interval for the population mean IQ score for **each** lead exposure group, separately. Summarize each confidence interval.

   b. Researchers were interested in assessing the difference in the mean IQ score between the two exposure group populations. Give the estimate mean difference, the standard error, and the 95% confidence interval for the **difference** in population mean IQ scores. Summarize the confidence interval.

   c. Researchers hypothesized that the exposed group (GROUP = 1) would have a lower population mean IQ score than the control group (GROUP = 2). Set up and conduct a statistical hypothesis test to address the research hypothesis. Carefully state the null and alternative hypotheses to be tested. Give the parameter of interest, the estimate of this parameter, the standard error of the estimate, the test statistic, and the p-value. Summarize the results of the test.