MSDS 6371 HW 2

1. The world’s smallest mammal is the bumblebee bat, also known as the Kitti’s hog nosed bat. Such bats are roughly the size of a large bumblebee! Listed below are weights (in grams) from a sample of these bats. Test the claim that these bats come from the same population having a mean weight equal to 1.8 g.

Sample: 1.7 1.6 1.5 2.0 2.3 1.6 1.6 1.8 1.5 1.7 1.2 1.4 1.6 1.6 1.6

* 1. Perform a complete analysis using SAS. Use the six step hypothesis test with a conclusion that includes a statistical conclusion, a confidence interval, and a scope of inference (as best as can be done with the information above … there are many correct answers given the vagueness of the description of the sampling mechanism).
  2. Inspect and run this R Code and compare the results (t statistic, p-value, and confidence interval) to those you found in SAS. To run the code, simply copy and paste the below code into R.

***sample = c(1.7, 1.6, 1.5, 2.0, 2.3, 1.6, 1.6, 1.8, 1.5, 1.7, 1.2, 1.4, 1.6, 1.6, 1.6)***

***t.test(x=sample, mu = 1.8, conf.int = "TRUE", alternative = "two.sided")***

1. In the United States, it is illegal to discriminate against people based on various attributes. One example is age. An active lawsuit, filed August 30, 2011, in the Los Angeles District Office is a case against the American Samoa Government for systematic age discrimination by preferentially firing older workers. Though the data and details are currently sealed, suppose that a random sample of the ages of fired and not fired people in the American Samoa Government are listed below:

**Fired**

34 37 37 38 41 42 43 44 44 45 45 45 46 48 49 53 53 54 54 55 56

**Not fired**

27 33 36 37 38 38 39 42 42 43 43 44 44 44 45 45 45 45 46 46 47 47 48 48 49 49 51 51 52 54

* + - * 1. Perform a permutation test to test the claim that there is age discrimination. Provide the Ho and Ha, the p-value, and full statistical conclusion, including the scope (inference on population and causal inference). Note: this was similar to an example in Live Session 1. You may start from scratch or use the sample code and PowerPoints from Live Session 1.
        2. Now run a two sample t-test appropriate for this scientific problem. (Use SAS.) *(Note: we may not have talked much about a two-sided versus a one-sided test. If you would like to read the discussion on pg. 44 (Statistical Sleuth), you can run a one-sided test if it seems appropriate. Otherwise, just run a two-sided test as in class.*  Be sure to include all six steps, a statistical conclusion, and scope of inference.
        3. Compare this p-value to the randomized p-value found in question 2a.
        4. The jury wants to see a range of plausible values for the difference in means between the fired and not fired groups. Provide them with a confidence interval for the difference of means and an interpretation.
        5. Given the sample standard deviations from SAS, calculate by hand :
    1. Pooled standard deviation (sp)
    2. The standard error of ()

1. Inspect and run this R Code and compare the results (t statistic, p-value, and confidence interval) to those you found in SAS. To run the code, simply copy and paste the code below into R.

***Fired = c(34, 37, 37, 38, 41, 42, 43, 44, 44, 45, 45, 45, 46, 48, 49, 53, 53, 54, 54, 55, 56)***

***Not\_fired = c(27, 33, 36, 37, 38, 38, 39, 42, 42, 43, 43, 44, 44, 44, 45, 45, 45, 45, 46, 46, 47, 47, 48, 48, 49, 49, 51, 51, 52, 54)***

***t.test(x = Fired, y = Not\_fired, conf.int = .95, var.equal = TRUE, alternative = "two.sided")***

1. In the last homework, it was mentioned that a Business Stats professor here at SMU polled his class and asked students how much money (cash) they had in their pockets at that very moment. The idea was that we wanted to see if there was evidence that those in charge of the vending machines should include the expensive bill / coin acceptor or if it should just have the credit card reader. However, another professor from Seattle University was asked to poll her class with the same question. Below are the results of the polls.

**SMU**

34, 1200, 23, 50, 60, 50, 0, 0, 30, 89, 0, 300, 400, 20, 10, 0

**Seattle U**

20, 10, 5, 0, 30, 50, 0, 100, 110, 0, 40, 10, 3, 0

* + - * 1. Run a two sample t-test to test if the mean amount of pocket cash from students at SMU is different than that of students from Seattle University. Write up a complete analysis: all 6 steps including a statistical conclusion and scope of inference (similar to the one from the PowerPoint). (This should include identifying the Ho and Ha as well as the p-value.) Also include the appropriate confidence interval. **FUTURE DATA SCIENTIST’S CHOICE!: YOU MAY USE SAS *OR* R TO DO THIS PROBLEM!**
        2. Compare the p-value from this test with the one you found from the permutation test from last week. Provide a short 2 to 3 sentence discussion on your thoughts as to why they are the same or different.

1. Consider the Samoan discrimination problem data (see data from question 2).

a. Calculate the estimate of the pooled standard deviation from the Samoan discrimination problem (see data from question 2). Use this estimate to build a power curve. Use a meandiff = 0.2. Assume we would like to be able to detect effect sizes between 0.5 and 2 and we would like to calculate the sample size required to have a test that has a power of .8. Simply cut and paste your power curve and SAS code. HINT: USE THE CODE FROM DR. McGEE’s lecture. Instead of using **groupstddevs**, use **stddev** since we are using the pooled estimate.

b. Now suppose we decided that we may be able to live with slightly less power if it means savings in sample size. Provide the same plot as above but this time calculate curves of sample size (y-axis) vs. effect size (.5 to 2) (x axis) for power = 0.8, 0.7, and 0.6. There should be three plots on your final plot. Simply cut and paste your power curve and SAS code. HINT: USE THE CODE FROM DR. McGEE’s lecture. Instead of using **groupstddevs**, use **stddev** since we are using the pooled estimate. The effect size here refers to a difference in means, though there are many effect size metrics, such a Cohen’s D.

c. Using similar code, estimate the savings in sample size from a test aimed at detecting an effect size of 0.8(meandiff) with a power of 80% versus a power of 60%.