Exercise 6 Micro 110 Spring

In this exercise, you will use Python to perform the analysis for the yield improvement exercise covered in slides 31 to 37 of Module 3.

The idea behind this exercise is to give an example of how one would actually perform the method described in those slides. What we are doing is testing the hypothesis that treatment B does something statistically significantly different than treatment A. The process we are dealing with is left undefined to show that it could be anything, but for the example, lets say that we are making windshields for Ford. Process A is the current process. We have a lot of data about process A. In process B, our engineer decided to add a different UV filter layer to the windshield.

In order to pass regulations, let's say that the windshield must prevent at least 90% of UV radiation from passing through. The data we collect is yield data. Each value represents the percentage of windshields that successfully blocked UV. Our engineer ran process A for 10 days and recorded the results. Then they ran process B for 10 days and recorded the results. We compare these two processes and see that process B has 1.3 higher yield on average.

It might be that this difference is not significant. Maybe if we took 10 days on process A and compared them to 10 other days on process A, we would see that the difference is 1.5 or 1.6. Maybe the yield just varies by a lot and process B happened to get lucky and look like it improved yield. This exercise describes how you might test whether process B got lucky or not.

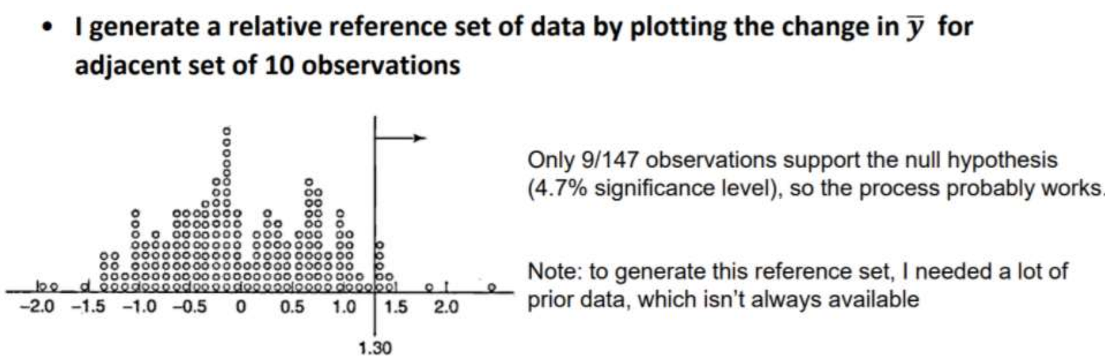
What we do is take our data from running process A since Ford started and evaluate it in the same way that we tested process B. We take 10 days of process A and compare the mean to the next 10 days of process A. We do this for the entire dataset we have on process A (just over 200 days). Then we see how many of those 10 day comparisons result in a change in the mean of greater than 1.3. If a change of greater than 1.3 happens all the time even when we don't change processes, it means that process B likely got lucky and doesn't actually help. If a change of greater than 1.3 barely ever happens, it means that process B is likely to be a real improvement and Ford should consider adopting it.

1) Import the text file Ex6data.txt, available for download from the Exercise 6 section of the Micro 110 moodle.

2) Generate a histogram of obs. Record the mean and standard deviation

3) You will now create a new array (or column if you're using excel), with the goal of matching the values in the ave10 column. Create a new array called CalcAve10, and assign it a formula to calculate the moving average of the past 10 observations. The values that you calculate should exactly match the values in ave10.

4) You will now create an array to generate a relative reference set of data by plotting the change in $y$bar for adjacent set of 10 observations. This process is demonstrated on slide 32 of module 2:



a. Calculate the difference in the moving averages.

b. Generate a histogram of this data. What is the mean and standard deviation for this histogram?

5) Now, suppose I perform a supposed process improvement, and then compare pre- and post-improvement sample means for sets of 10 adjacent samples. I find that the change in sample means

is 1.30, as shown below:

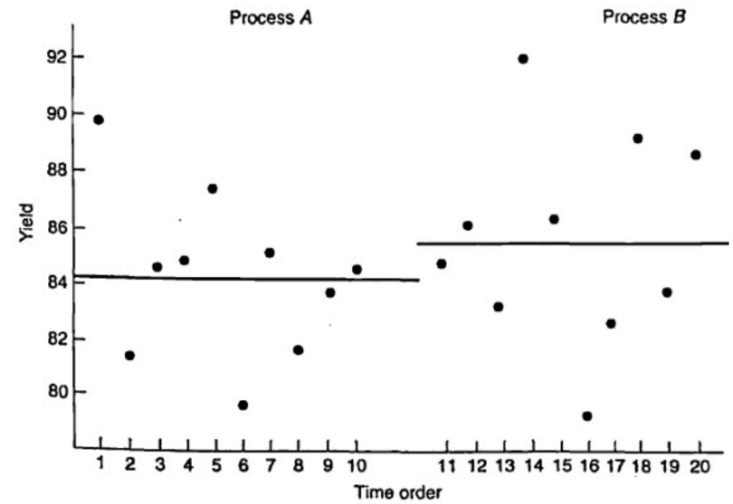


Table 3.1. Yield Data from an Industrial Experiment

Time Order	Method	Yield
1	*A*	89.7
2	*A*	81.4
3	*A*	84.5
4	*A*	84.8
5	*A*	87.3
6	*A*	79.7
7	*A*	85.1
8	*A*	81.7
9	*A*	83.7
10	*A*	84.5
11	*B*	84.7
12	*B*	86.1
13	*B*	83.2
14	*B*	91.9
15	*B*	86.3
16	*B*	79.3
17	*B*	82.6
18	*B*	89.1
19	*B*	83.7
20	*B*	88.5

$\bar{y}_A = 84.24$ $\bar{y}_B = 85.54$
$\bar{y}_A - \bar{y}_B = 1.30$

a) What is the probability that the process B actually improved things? HINT: You should determine the probability that the reference data from part 4 above could actually result in a sample mean change for sets of 10 adjacent samples of 1.30. Perform this test by comparing against the empirical data from part 4 and also comparing against a normal distribution with the mean and standard deviation of part 4.