Implementing SAS Hypothesis Testing

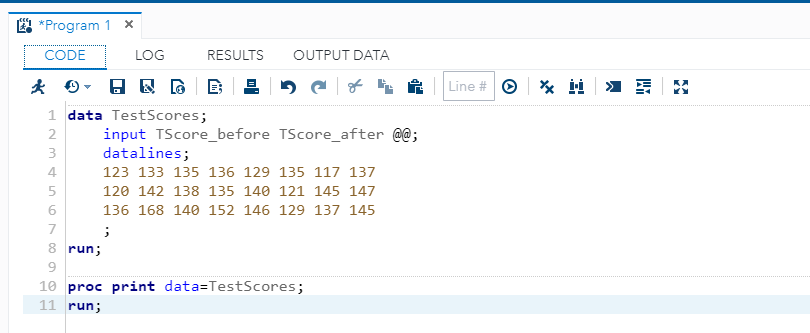
Tom Hefner

Colorado State University – Global Campus

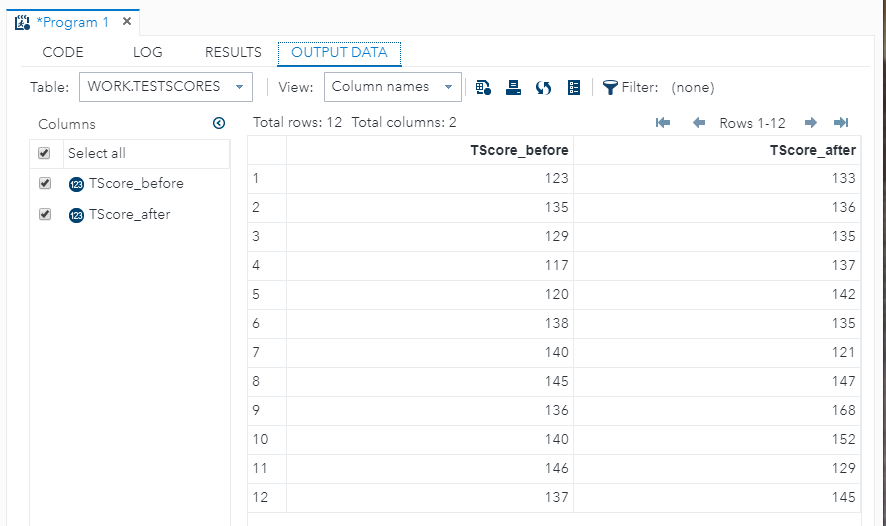
For this test I will be using SAS Studio to perform a T-Test on the hypothesis regarding tutoring and test scores. More specifically we will test the impact of tutoring on test scores for 12 individuals. We will test the hypothesis that tutoring does not have an impact with alternate hypotheses that tutoring can either have a positive or negative impact.

The first step in performing a hypothesis test on data is to establish a dataset. In the following code we are telling SAS that our data will be stored in table TestScores with headers TScore\_before and TScore\_after to differentiate the test scores.

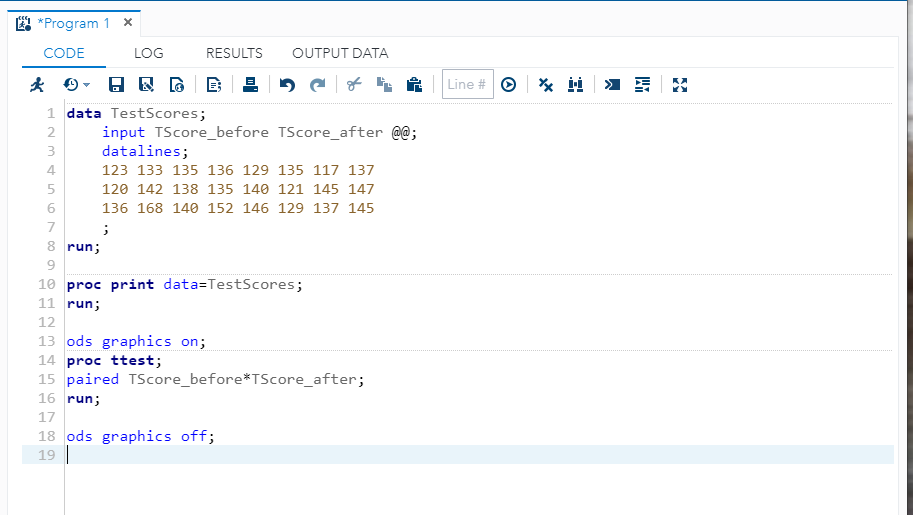
Do note that space bar executions are required between figures as tab spacing omits data points.



The execution of the above code provides us a simple two column 12 observation table as seen below. In this screenshot we are also verifying that the prior code executed properly. It’s key to have the code display properly as misaligned data would render any given results in error. SAS was coded to have 12 observations in 2 columns and we received just that below.



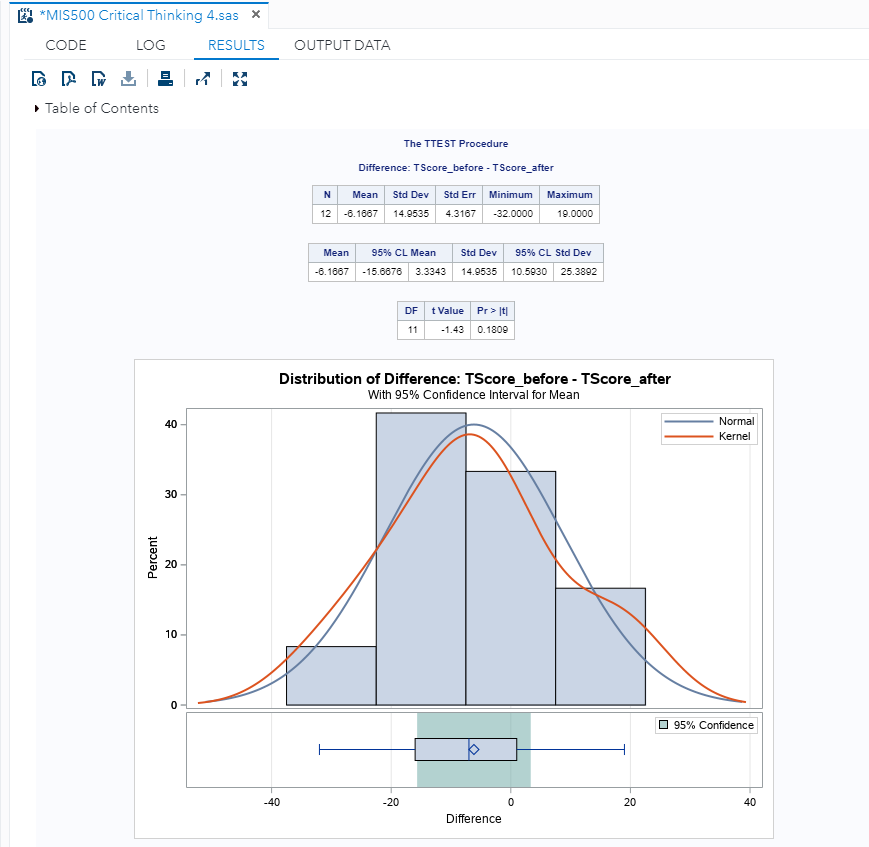
The next step covers two parts in our hypothesis testing. The first step to discuss is the command to tell SAS to enable graphical representations of our dataset TestScores. Furthermore, the enable graphics are to be performed according to proc ttest on both of our columns of data. In very little code we are actually asking SAS to perform multiple procedures all at the same time.



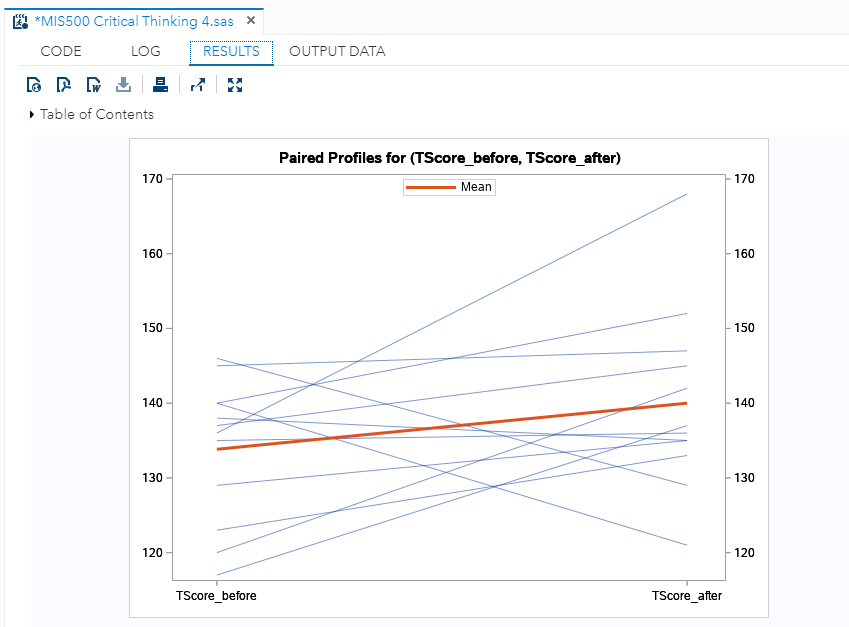
With the successful execution of our SAS code we have a number of screenshots to review. This first screenshot is identical to our output screenshot above with the exception that it has a more polished look. The key here is that our data is intact and is laid out how we expect it.



The following screenshot is our first true look at the data results. We are first given the statistical references of the data followed by a bell curve distribution. In looking at our box plot subset below we observe that the overall mean performance increased. In looking at the scores compared to a 95% bell curve we can also see these results as our kernel curve in a number of observations falls above the normal distribution. Furthermore, if we continue to look at the bell curve we can observe that those scores at both ends of the bell curve appear to have been more positively affected by tutoring while those scores within 1 standard deviation did not benefit.

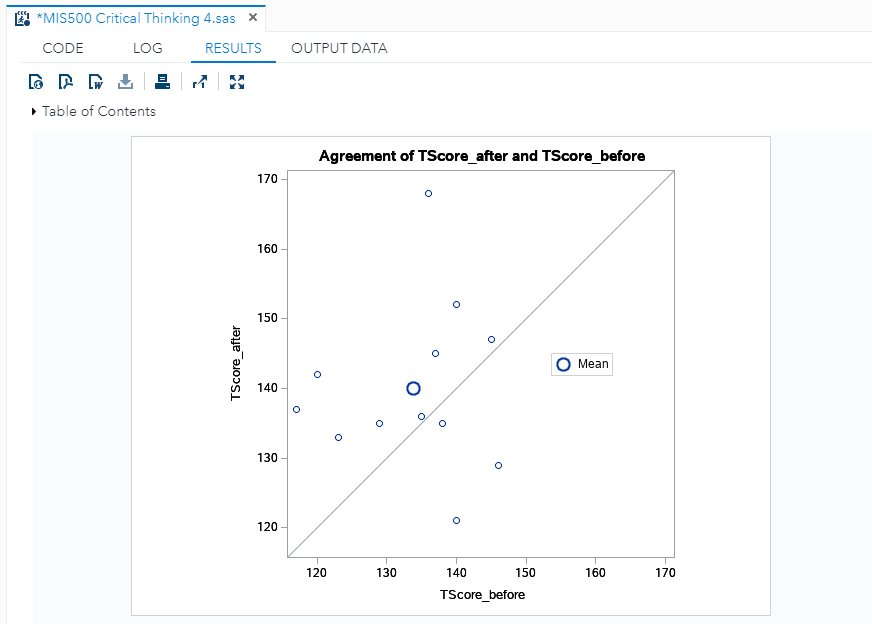


In our next slide we are looking at the before and after scores for each of the 12 observations. Our first observation is the overall mean for all test score comparison shows an increased mean. Although one might argue this is an acceptance of the null hypothesis that tutoring has a positive impact, we need to note that not all scores improved. In fact, some scores decreases significantly as indicated by the sharp decline from top left to bottom right.

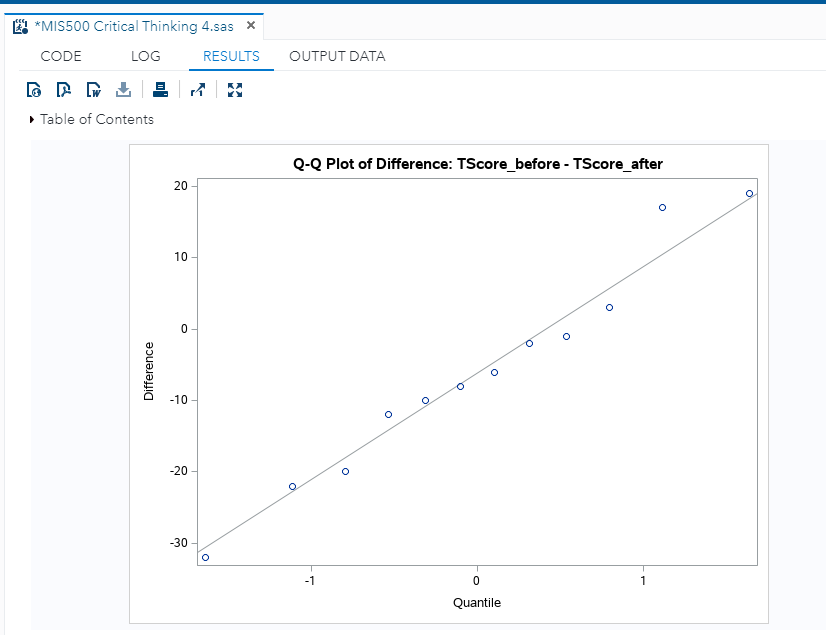


Below we continue to look at our 12 observations and the potential for tutoring impact. We can see that 9 of the 12 test score comparisons fall above our line indicating that the scores of 9 students improved while 3 declined.

We are starting to see that there is a tutoring impact. Each of our graphs thus far indicate both positive and negative movement in test scores.



In our last slide we review a Q-Q plot. In our Q-Q plot we have our normal distribution as displayed by the line through the graph. As another validation of our data we see some variations of observations that do not deviate much from our line. With the exception of one data point in the top right corner our observations are not too far in deviating from a normal distribution.



The null hypothesis that tutoring has no effect on tutoring is not accepted. In our 12 observations we saw a varying degree of performance by 12 students in that tutoring provided both an increase and decrease in student performance. Student performance increases more for those students with the higher and lower before test scores. Individuals with scores within approximately 1 standard deviation were not impacted to the same degree. Some student scores increased while other decreased. Having reviewed the data we confidently reject the null hypothesis and accept the alternate hypotheses.