Data Analyst Nanodegree Assignment I—The Stroop Task

Question 1: What is our independent variable? What is our dependent variable?

Answer: The independent variable is the word color congruency, i.e., whether or not the colors in which the words are printed match the names of the colors. The dependent variable is the time taken to read the names of each of the colors in each task.

Question 2: What is an appropriate set of hypotheses for this task? What kind of statistical test do you expect to perform? Justify your choices.

Answer: The null hypothesis for this task is "The time for the congruent words condition will be the same as the time for the incongruent words condition." Mathematically, the null hypothesis is stated $H_0: t_1 = t_2$, where H_0 is the null hypothesis, t_1 is the time for the congruent words task, and t_2 is the time for the incongruent words task. The alternative hypothesis is "The time for the incongruent words condition will be more than the time for the congruent words condition." Its mathematical statement is $H_1: t_2 > t_1$, where H_1 is the alternative hypothesis, t_2 is the time for the incongruent words task, and t_1 is the time for the congruent words task. I expect to perform Student's t-test for dependent means because the samples are random (there is no way to predict the times in either task), the task has been repeated several times, and the scores are approximately normally distributed.

Question 3: Report some descriptive statistics regarding this dataset. Include at least one measure of central tendency and at least one measure of variability.

Answer: The mean for the times in the congruent task is approximately 14.05 with a standard deviation of 3.56, and the mean for the times in the incongruent task is approximately 22.02 with a standard deviation of 4.8. The range for the times in the congruent task is 13.698 and the range for the times in the incongruent task is 19.568.

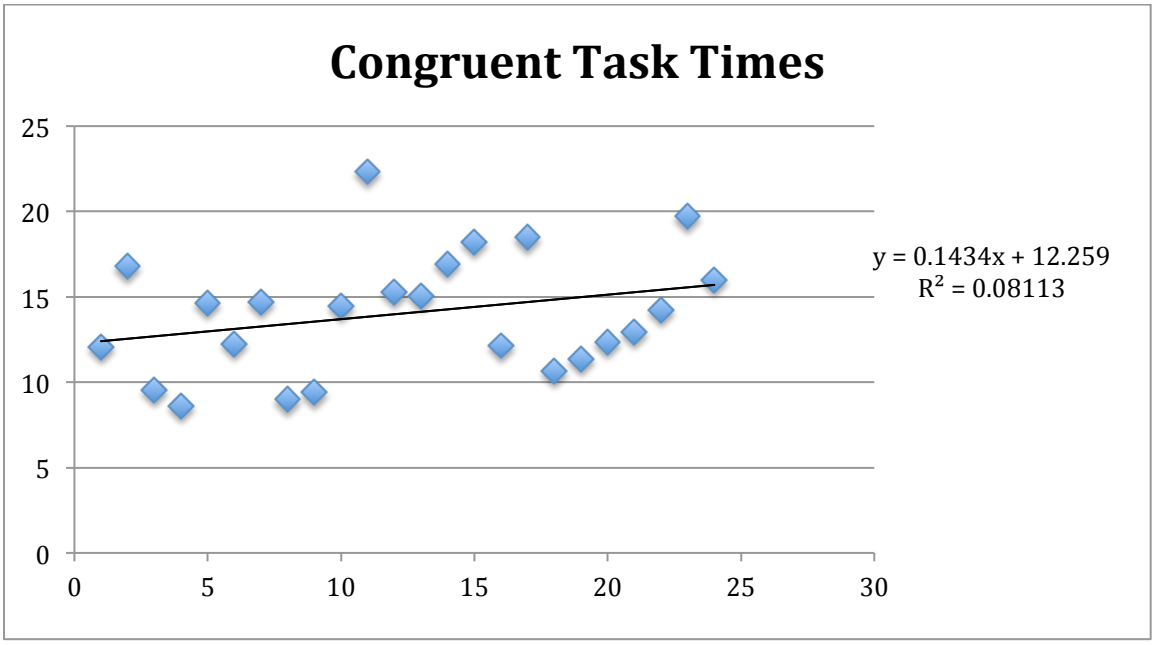
Question 4: Provide one or two visualizations that show the distribution of the sample data. Write one or two sentences noting what you observe about the plot or plots.

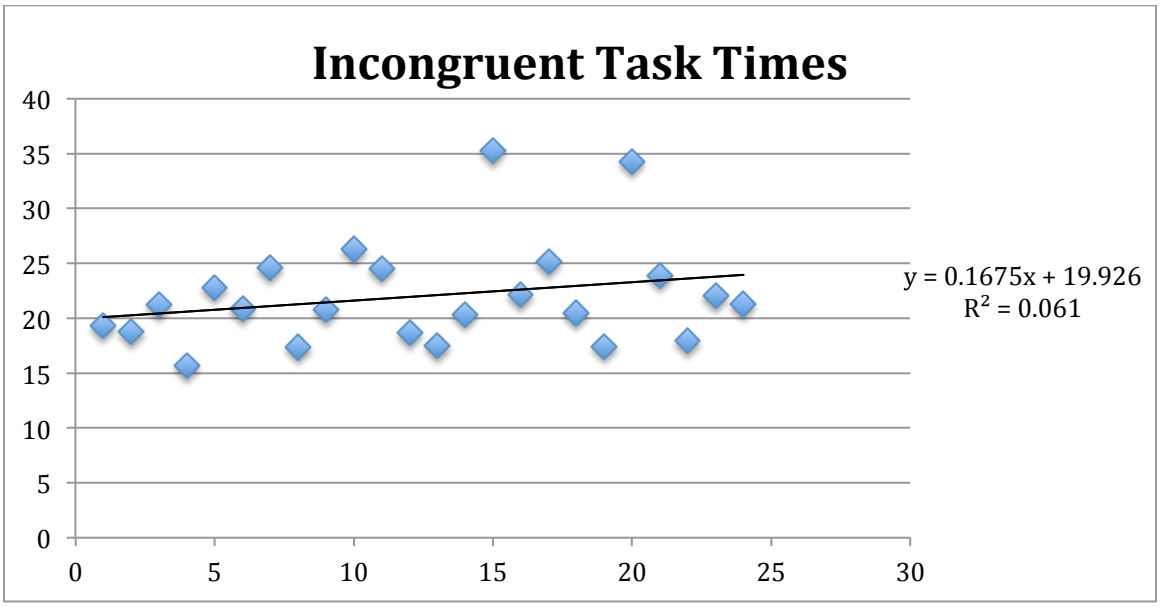
Answer: After plotting the times for the congruent task and for the incongruent task, I noticed that even though the times for the incongruent task appear much closer together than those in the congruent task, the former has a lower r^2 value than the latter. See the following page for the plots.

Question 5: Now, perform the statistical test and report your results. What are your confidence level and your critical statistic value? Do you reject the null hypothesis or fail to reject it? Come to a conclusion in terms of the experiment task. Did the results match up with your expectations?

Answer: The mean of the times in the congruent task is 14.05 and the mean of the times in the incongruent task is 22.01, so the mean difference is -7.96. The standard deviation

of the mean difference is 0.99. The standard error of the mean difference is $\frac{0.99}{\sqrt{24}}$, or 0.202 which means that if a 95% confidence level is assumed, the t-statistic is 8.02, and under the one-tail test, the p-value is 0.00001, so we fail to reject the null hypothesis.





List of websites used for this assignment:

1. https://support.office.com/en-us/article/TTEST-function-1696ffc1-4811-40fd-9d13-a0eaad83c7ae
2. http://blog.minitab.com/blog/adventures-in-statistics/regression-analysis-how-do-i-interpret-r-squared-and-assess-the-goodness-of-fit
3. https://en.wikipedia.org/wiki/Central_tendency

4. http://stattrek.com/descriptive-statistics/variability.aspx?Tutorial=AP