Stroop Effect Data Analysis

Variables, Hypotheses, Statstical Test:

The data for this example are the times to complete the "Stroop Test" for each subject in the congruent and incongruent conditions. The condition under which the test is administered (congruent versus incongruent) is manipulated by the investigator and is therefore, the independent variable. The time to completion of each trial is being measured for each subject in each condition, and is the dependent variable. An intuitive prediction about the task is that it would be more difficult when the colors and words are incongruent.

Let μ_c represent the population mean for the congruent condition, and μ_i represent the population mean for the incongruent condition. If we predict the incongruent task is more difficult, null (H_o) and alternative (H_a) hypotheses can be stated as follows:

H_o: $\mu_i \leq \mu_c$

H_a: $\mu_i > \mu_c$

Let μ_d be the difference between the means defined as $\mu_d = \mu_i - \mu_c$. We can then state these hypotheses using the difference between the means as follows:

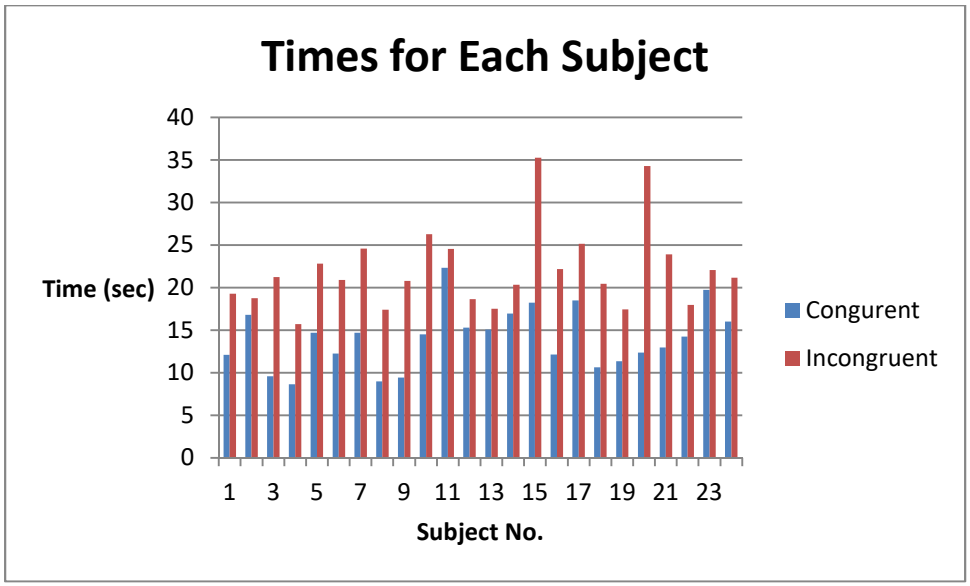
H_o: $\mu_d \leq 0$

H_a: $\mu_d > 0$

The data from the study allow us to calculate the sample means for the congruent (M_c) and incongruent (M_i) conditions. The mean difference, which is the same as the difference between the means, is $M_d = M_i - M_c$. Because we do not know the population mean or variance, we need to use the t-test instead of the z-test of significance. Since the study involves testing two different conditions for the same subject, the t statistic for correlated samples**,** rather than independent samples, would be the appropriate statistic to use for testing significance of the results.

Results: Summary and Visualization:

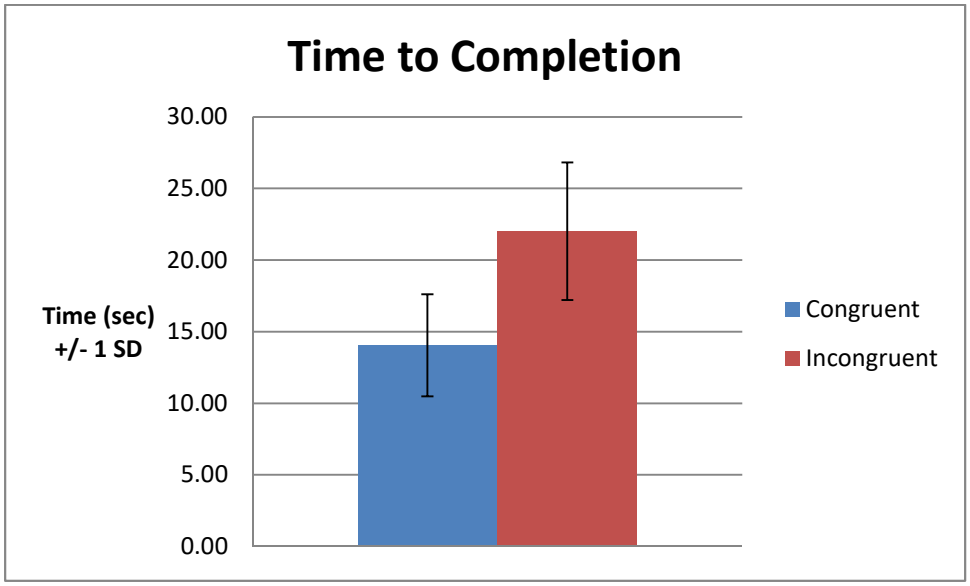
This chart shows the time to completion in each condition for each subject:



It can be seen easily from this chart that the time to complete the task is longer for the incongruent than the congruent condition in every subject

The mean, range, and standard deviation for each condition are as follows:

Condition	Mean (range)	Std.Dev.
Congruent	14.05 (8.63-22.330	3.56
Incongruent	22.02 (15.69-35.26)	4.80

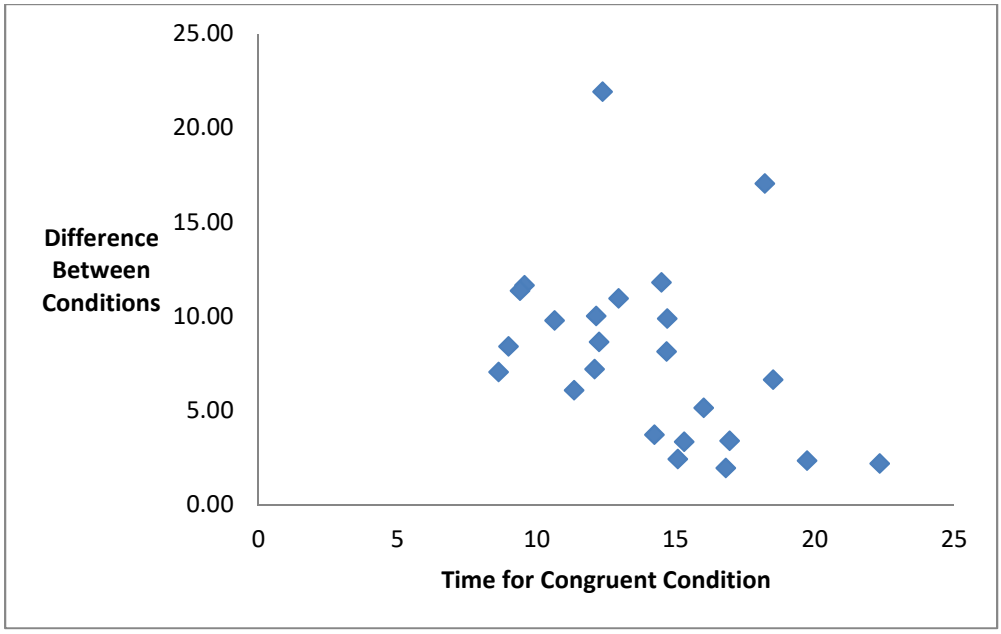


Significance Test and Probability

The mean difference (Md) for the two conditions is 7.96 seconds. Using this as the point prediction for the sampling distribution the standard error (SE) for the sampling distribution is 4.86, and the t-statistic for this sample is 8.02. For a one-tailed test (df = 23) this is significant at a level of $p < .0001$. The t-critical value for a one-tailed test, df=23, at a 99% confidence level is 2.500. A 99% confidence interval for the mean difference between the two conditions is 4.78-11.15 seconds. The results show a high level of statistical significance and therefore the null hypothesis should be rejcted. It appears that having the color of the word presented being different from the actual word makes the task much more difficult.

Discussion

One striking aspect of this data set, aside from the strength of the effect, is the variability across subjects. This variability suggests a way to examine one hypothesis about the reason for the Stroop effect. A possible explanation for the effect is that reading the words is easier than naming the colors. If this is the case, faster readers might show a greater difference between the congruent and incongruent conditions on the test. We could use the time to complete the congruent task as a proxy for reading speed and see if this is correlated with the difference between the two conditions. The chart shows a scatter plot of these two variables for each subject. If one excludes the two points which appear to be outliers, it does suggest a correlation.



On the other hand, it could be that reading any word slows down color recognition, which could be tested using words not describing colors and nonsense series of letters of similar size as the conditions.

An interesting phenomenon!

Resources used in completing this project:

1. Udacity inferential statistics course for review of the necessary concepts.
2. Vassarstats website statistical calculator, to check my calculation of the t-statistic.
3. Microsoft Excel for calculating means, etc. and preparing the charts.