Project 2
Statistics for Stroop task

1. The independent variable is the list of words, with each word displayed in a color of ink. There are only two conditions for this variable, congruent (word matches its color) or incongruent (word not matching its color).
The dependent variable is the time T required by a subject to name the ink colors in the equally-sized lists.

2.
Hypothesis
The null hypothesis H0:
$dT \leq 0$
i.e. The paired difference in recitation time (incongruent time minus the congruent time for a given subject), when averaged across the entire population of subjects, is less than or equal to zero.
The alternative hypothesis HA:
$dT > 0$
i.e. The paired difference in recitation time (incongruent time minus the congruent time), when averaged across the entire population of subjects, is greater than zero.

where
$dT = Tinc - Tcon$ where each is a time measure for the population (all English-literate humans).
Tinc = mean time required by population to recite the incongruent word list.
Tcon = mean time required by population to recite the congruent word list.

Statistical Test
The test I choose is the one-tailed, paired-sample t test.
I choose t (rather Z) test because the N=24 which is small (<30).

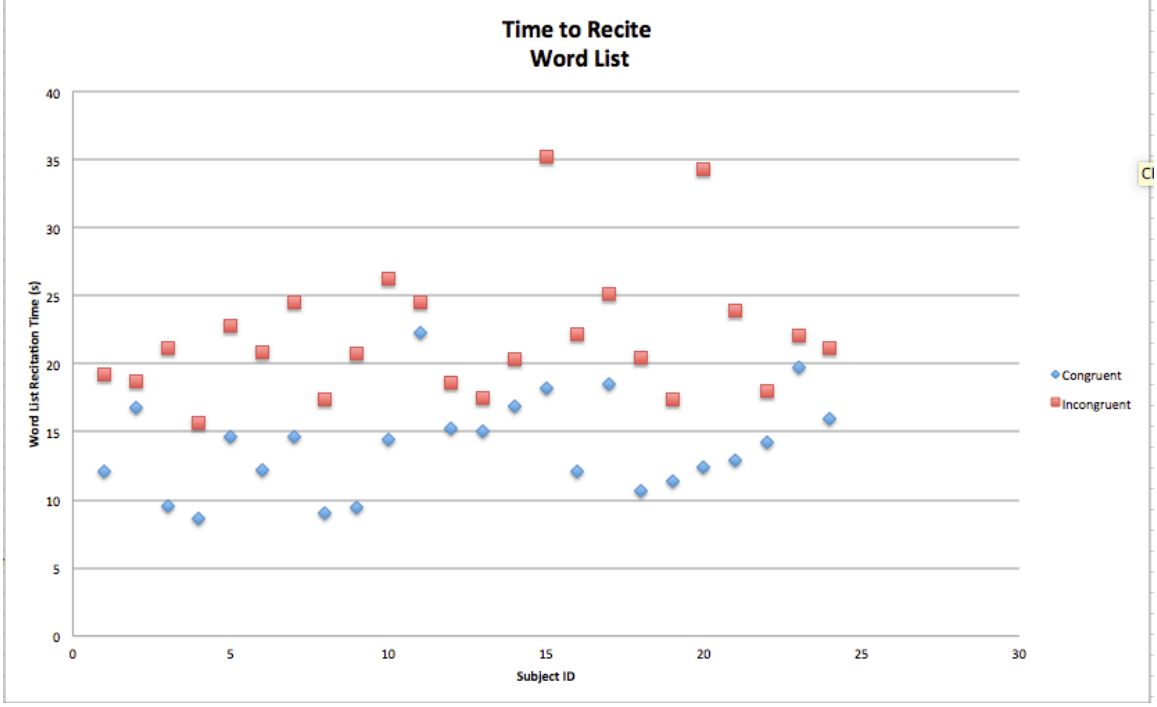
The reason for choosing a hypothesis that is one-tailed ($dT \leq 0$) instead of two-tailed ($dT = 0$) is the desire to test the expected (intuitive) effect that it will take more time for the subject's brain to process and speak the color in the incongruent condition. Therefore the null hypothesis encompasses all other results, including the possibility that $dT < 0$.

The reason for choosing a "paired test" is that we want to control for the individual abilities of the subjects to perform the test and focus the test on the effect of congruency. Therefore we look at differences in recitation time, paired by subject, instead of treating the congruent and incongruent data as independent samples.

3. For the **congruent condition**
a) Central tendency measured by average and median
Average time: 14.05 s
Median time: 14.36 s
b) Variability
Standard deviation (sample): 3.56 s
Range: 8.63 to 22.33 s

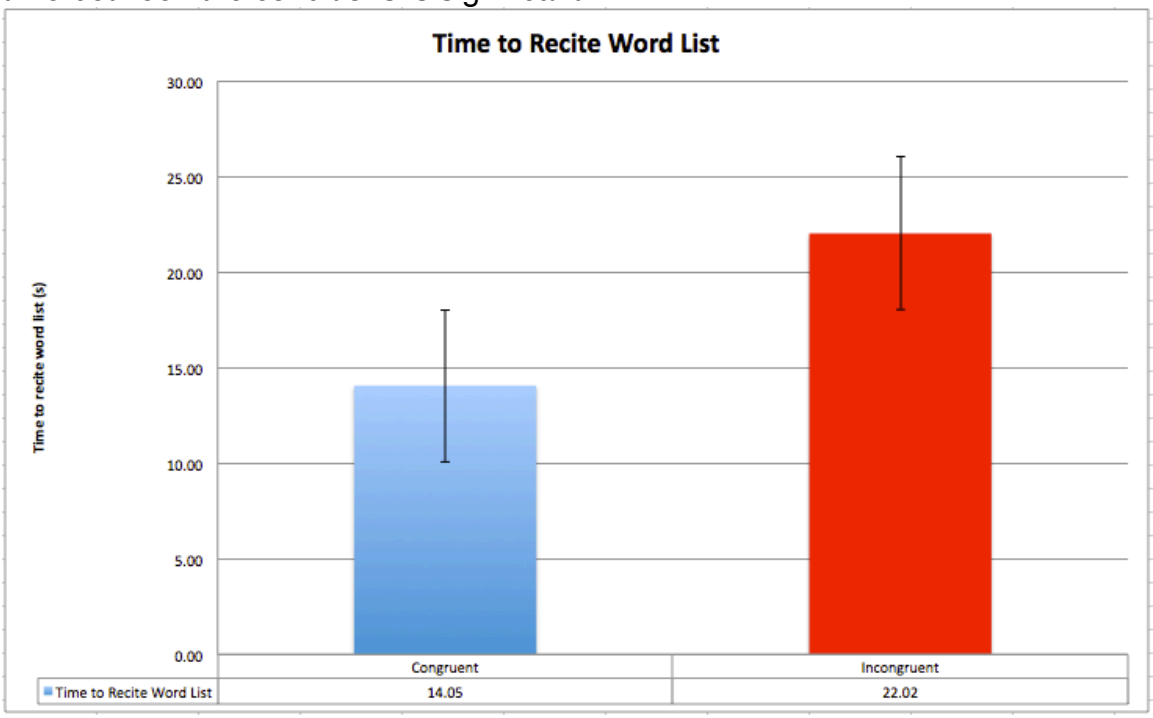
For the **incongruent condition**
a) Average time: 22.02 s
Median time: 21.02 s
b) Standard deviation (sample): 4.80
Range: 15.69 to 35.26 s

4. I've created two charts:
a) Chart 1 is a scatter plot showing the Time to Recite Word list as a function of the subjects, with Red marking incongruent times, Blue marking congruent times. This gives me the general impression that congruent times are shorter, and that there is no large skew present (i.e. we should be able to use statistics appropriate for a normal distribution.) Moreover, the red square is above the blue square for every single subject in our sample, which indicates that regardless of ability at this task, subjects tend to be faster at the congruent task.



b) Chart 2 is a bar chart showing the mean Time across subjects for the 2 conditions. The error bars are the standard error reported by Excel. The error bar from the Congruent condition has very little overlap with the error bar from the

Incongruent condition, which leads me to believe that the difference in recitation time between the conditions is significant.



5. The statistical test I choose is the paired t-test, one-tailed, with the assumption of equal variances.
NOTE: Details of calculations are in submitted Excel worksheet.
The critical alpha is p=0.01.
The critical t-statistic is 2.500.
The t-statistic for this data is 8.02.
The probability associated with this paired t-test (one-tailed) is 2.05 x 10^-8 (< alpha=0.01).
The t0 > tcrit therefore **I reject the null hypothesis**, which means that HA is accepted: the time required for the incongruent data is significantly longer.
The results do match my expectations considering the clear effect shown in the scatterplot.

6. I would speculate that the effect is caused because the brain requires more processing time to resolve the conflict due to incongruency. There may be two networks, one for processing color and one for processing word meaning. A third decision-making process must be engaged to determine which “answer” (color or word meaning) is to be selected for recitation.

A similar task would be to provide the subject with a series of images of objects (e.g. car, truck, apple, etc.) with a word label on each. The congruent condition would have the word label matching the object, and non-matching in the incongruent case. This would be interesting to see if the effect is similar if the color is replaced by object type.

Reference:

Biostatistical Analysis by Jerrold H. Zar, 2nd ed., 1984.