Project Specification

1) Question response correctly identifies the independent and dependent variables in the experiment.

   - **Independent Variable** is the word condition (word congruency); either being congruent or incongruent. Its strength (measured in pounds).
   - **Dependent Variable** is the amount of time it takes to name, out loud the color of the links in which the words are displayed. Its spend time (measured in seconds).

2) **A) Null and alternative hypotheses are clearly stated in words and mathematically. Symbols in the mathematical statement are defined.**

- H0: There would be a lower mean time for the incongruent tests than the congruent tests or no noticeable difference in time duration between the congruent and incongruent tests
- HA: There would be a noticeable increase in time duration between the congruent and incongruent tests. The experiment should be a single-tailed type scenario as there the question does mention "...does incongruence increase response times" as well as the incongruent times were predominantly larger than the congruent times.
- **μ1** = the mean of time spent for the participant group of congruent test
- **μ2** = the mean of time spent for the participant group of incongruent test
- **The null hypothesis (H0): μ1 - μ2 ≥ 0**, the time duration difference of the congruent and incongruent tests would equal 0 or be greater than 0, meaning the incongruent times would be less or equal to the congruent times for the population.
- **The alternative hypothesis (HA): μ1 - μ2 < 0**, the time duration difference of the congruent and incongruent tests would be less than 0, meaning the incongruent times would be greater than the congruent test times for the population.
- **You are right the main idea of this test is that we are using a sample in order to make inferences about the population that is unknown, to find whether there is a difference between congruent and incongruent means in the population.**
- **There is no difference in population means of response time under incongruent and congruent scenarios which can be mathematically represented as: To meet this specification I need a clear definition of the null hypothesis, which must states that there will be no change in the means or as the example bellow:**
  H0= μc = μI

b) **A statistical test is proposed which will distinguish the proposed hypotheses. Any assumptions made by the statistical test are addressed.**

- **Dependent**: I chose to go the dependent t-test route as even though we do two tests (congruent and incongruent) it is with the same group of people that was tested on the congruent words then tested on the incongruent words.

• **Paired**: I chose to go the paired t-test because the participants measured at two time point (congruent and incongruent word sets).

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

Central Tendency: Going into the data set of the Stroop experiment determining.
The **means** of the congruent recordings (designated xc) and the incongruent (designated xi)

x Mean c = 14.05 x Mean i = 22.02

The **median** of the congruent group (designated as **C med**) and the incongruent group (Designated as **I med**)

C med = 14.36 I med = 21.02

Considering all data on the congruent and incongruent test groups have unique time no same time as the data is continuous there is no mode available on both test groups.

Measure of Variability: To get a better understanding of the general variability of the data sets (congruent and incongruent) we can find the IQR which accounts for 50% of the data for the perspective set of data.

First arrange the data values in order of least to most and then find the values of the 1st quartile and the 3rd quartile:

Cq1 = 16.201 Iq1 = 24.052 Cq3 = 11.895 Iq3= 18.717

After finding the values of the first and third quartile we can determine the inner quartile range (IQR) for each data set.

C IQR = 4.306 I IQR =5.335

Considering the variability of each test group we see when ordering the values and finding the range we get (C range designated as range of congruent sample and I range as range of incongruent sample):

C range = 13.70 I range = 19.57

Average Deviation for the two test groups (congruent and incongruent) can be calculated with the Known mean of x mean c = 14.05 and x mean i = 22.02 using this formula:

Average Deviation: $\frac{\text{Sum } |Xi - \text{Mean}|}{N}$

C (ave.dev) = 2.85 and I (ave.dev) = 3.40

Next measure we want to calculate is the measurement Variance with the data sets given (Congruent and Incongruent Groups)

To find the Variance of the data sets however we first need to determine the Square of Sums so we can have a simpler Variance equation.

$\sum i$=24 ($xi - x\bar{x}$) ^2)
(Sum of Squares c) SSc = 291.388 (sum of squares i) SSi = 529.270

Given: SS (Square of Sums) as like numerator and the denominator of n-1, n represent the participants on a given test (n = 24):

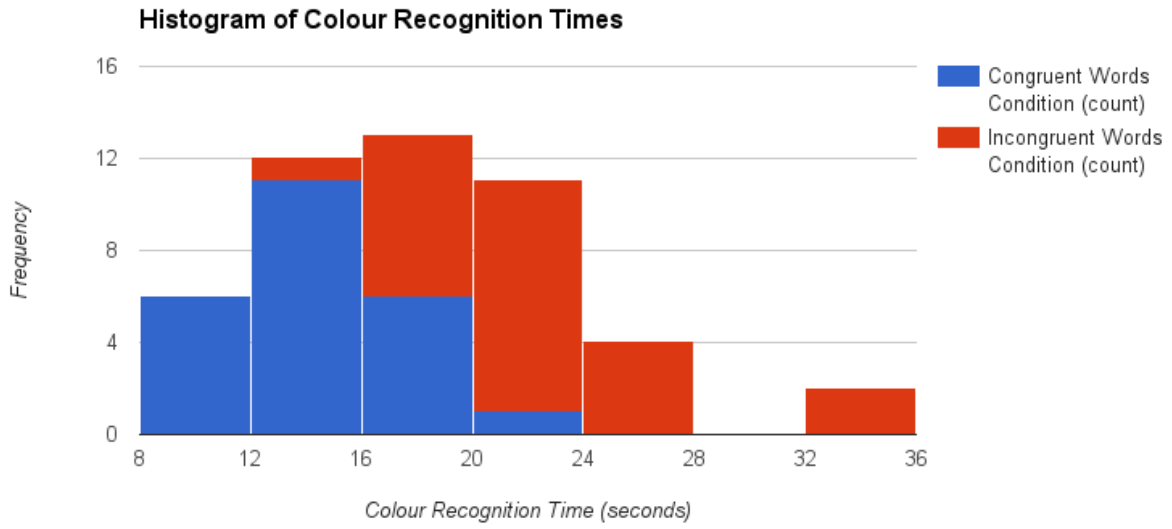
$$Var(C) = \frac{}{n-1}$$

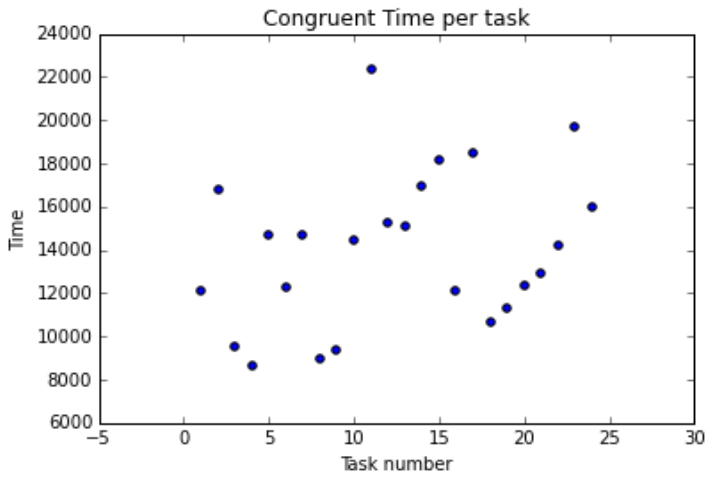
Therefore, **Var (congruent) = 12.669 and Var (incongruent) = 23.012**

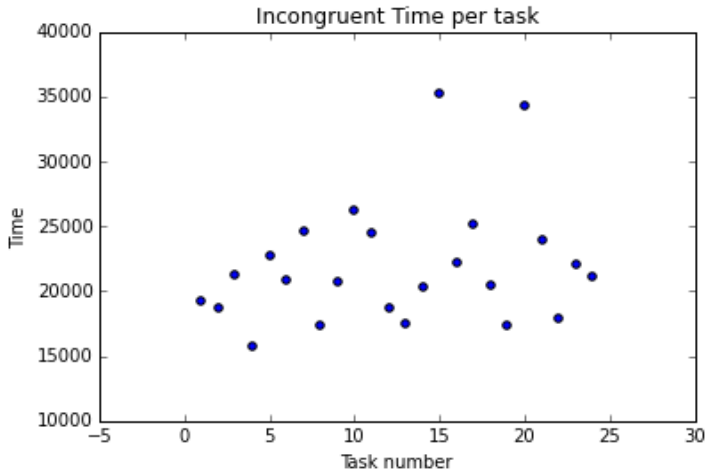
Sample standard deviation for each of the test groups (**congruent and incongruent**).
Sample Standard Deviation = square root of variance
Sample SD (congruent) = 3.559 Sample SD (incongruent) = 4.797

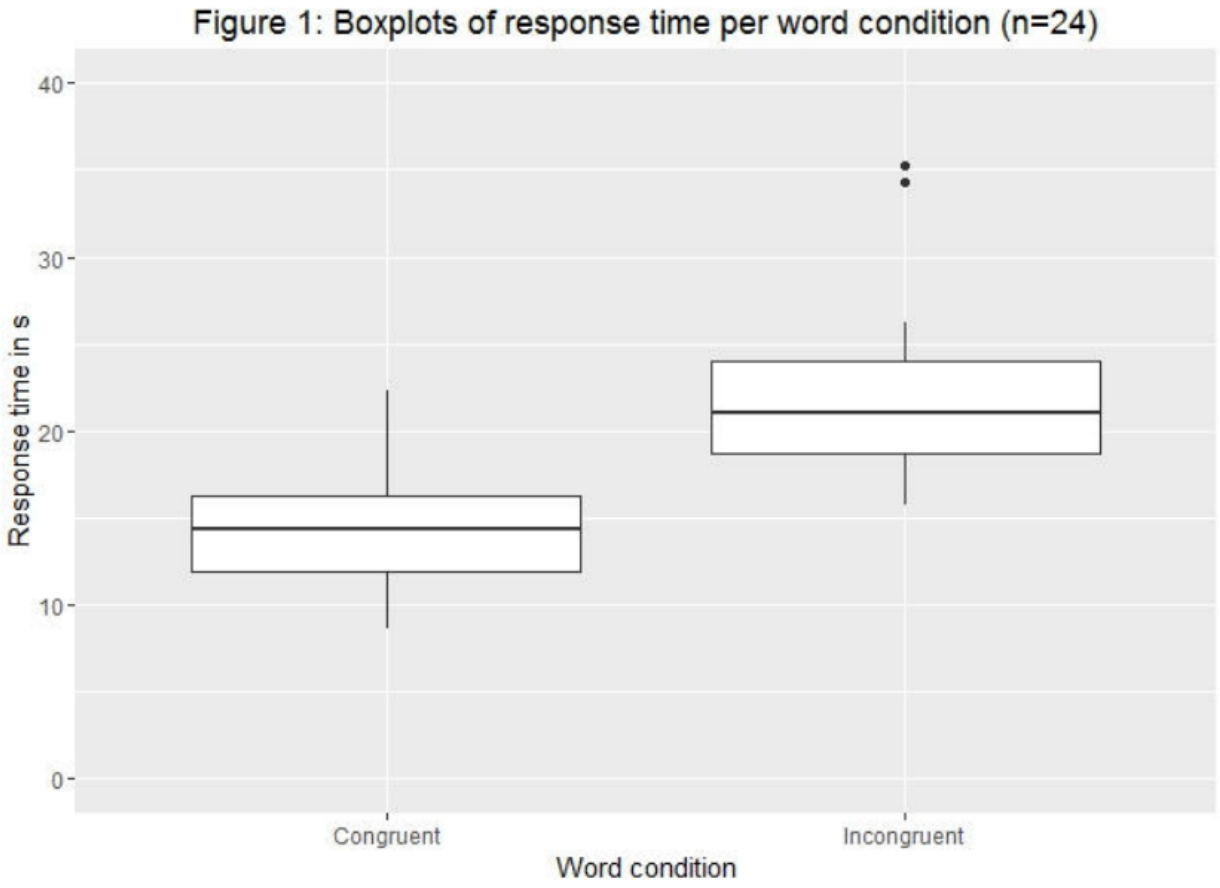
4.) **One or two visualizations have been created that show off the data, including comments on what can be observed in the plot or plots.**

- The distribution of the sample data can be visualized with the boxplot shown below. Based on the plot, the reaction time of the task (RTT) under incongruent words condition appears to be longer than that of the congruent words condition, which is consistent with the sample statistics generated earlier. Also, the distribution of the sample data from the incongruent group is likely to be positively skewed due to two outliners.




Congruent Time per task
Time
Task number
24000
22000
20000
18000
16000
14000
12000
10000
8000
6000
−5
0
5
10
15
20
25
30


Incongruent Time per task
Time
Task number
40000
35000
30000
25000
20000
15000
10000
−5
0
5
10
15
20
25
30



5) **A statistical test has been correctly performed and reported, including test statistic, p-value, and test result. The test results are interpreted in terms of the experimental task performed**

* We got from question # 3 $\bar{x}$ **Mean c = 14.05** $\bar{x}$ **Mean i = 22.02**
  Then we can calculate a point of estimate by using $\bar{x}$ **Mean c -- ** $\bar{x}$ **Mean i**
  $\bar{x}$ **Mean c --- ** $\bar{x}$ **Mean i**
  14.05 --- 22.02 = -7.97

Using the average of the differences which should match the point estimate of -7.97 we subtract that value from the differences of each person and then square it:
(Difference of Congruent and Incongruent – Average of differences)^2
Having this set of squared deviations from the mean we can sum them up into
Square of Sums Using $\sum_{i=24} (x_i - \bar{x}_{mean})$
Square of Sums of difference of sample means = 544.3304

Given the SS (Square of Sums) we can use this to determine the variance of the difference. Since we are dealing with a sample we need to use the equation with the denominator of n-1, n being the number of participants on the given tests (n = 24)

$$\text{Variance of difference} = \frac{\text{square of sums of difference of sample mean}}{n-1}$$

Variance of difference = 23.667

Standard Deviation = Square root of Variance
= 4.865

Now, we can calculate the t-statistics.

$$\text{T-statistic} = \frac{\boldsymbol{x}\text{mean}\boldsymbol{C} - \boldsymbol{x}\text{mean}\boldsymbol{i}}{\boldsymbol{ss}/\text{square root of n}} = \frac{-7.97}{4.865/\text{square root of 24}}$$

$$= -8.026$$

Then, we gone calculate the T-Critical Value

n=24, the degree of freedom for this question is n-1 = 23. Since, we are aiming to find with 99% Confidence level for a one tail our $\alpha = 0.005$

Given df = 23 and $\alpha = 0.005$ @ 99% Confidence Level: need to be between +- 2.807 Given that the T-Statistic found was - 8.026 it leads us to **reject the null hypothesis** due to T-Statistic being in the critical area of tested at 99% Confidence Level.

Therefore, the t-test confirms what was expected that incongruent tests would normally take longer than congruent tests as congruent test linked both visual colors with the correct printed name of the color while incongruent required to try to separate the visual color and the name of the color which required a bit more time to do.

I refer Data Analysis for Mangers (DS 101) California State University Text Book and I used a note from Udacity lecture. And also I refer the website: examples.yourdictionary.com/independent-and-dependent-variance-examples.html