Introduction

As part of this project, a test has to be performed to analyze the reading time for a list of words in given two conditions. In the experiment, participants are presented with a list of words, with each word displayed in a color of ink. The participant's task is to say out loud the color of the ink in which the word is printed in the following two conditions:

```
1.     Congruent words condition – In this condition, the names of the word
match the colors in which they are printed: for example RED, BLUE.
2.      Incongruent words condition – In this condition, the words displayed
are color words whose names do not match the colors in which they are prin
ted: for example PURPLE, ORANGE
```

In each case, we measure the time it takes to name the ink colors in equally-sized lists. Each participant will go through and record a time from each condition.

In a Stroop task, participants are presented with a list of words, with each word displayed in a color of ink. The participant's task is to say out loud the color of the ink in which the word is printed. The task has two conditions: a congruent words condition, and an incongruent words condition. In the congruent words condition, the words being displayed are color words whose names match the colors in which they are printed: for example RED, BLUE. In the incongruent words condition, the words displayed are color words whose names do not match the colors in which they are printed: for example PURPLE, ORANGE. In each case, we measure the time it takes to name the ink colors in equally-sized lists. Each participant will go through and record a time from each condition.

Questions For Investigation

1. What is our independent variable? What is our dependent variable?

In this experiment, Independent variable are the two conditions :

```
1. Congruent words condition where the names of the word match the colors
in which they are printed: for example RED, BLUE.
2. Incongruent words condition where the words displayed are color words w
hose names do not match the colors in which they are printed: for example
PURPLE, ORANGE
```

Dependent Variable is the time taken by the participants to name the ink colors in equally-sized words lists in each condition

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

In this experiment, the Null Hypothesis is that the time taken to read the list of colors in incongruent condition will be less than or equal to the time taken to read the list of colors in congruent condition.

$$\mu_i <= \mu_c$$

or

$$\mu_i - \mu_c <= 0$$
$$\mu_d <= 0$$

And Alternative Hypothesis will be that, the total average time taken to read the colors in incongruent condition will be more than the total average time taken to read the list of colors in congruent condition.

$$\mu_i > \mu_c$$

or

$$\mu_i - \mu_c > 0$$

Where
μ_c -> population mean time in congruent condition,
μ_i -> population mean time in incongruent condition

Here the data will be analyzed using One-tailed T-test in positive direction for dependent variables.

T-test - because the population parameters are not known

One-tailed test in the positive direction- as we need to check if the sample mean for the incongruent condition is significantly higher than the sample mean for the congruent condition.

Dependent - Because the same subject is taking test twice or put to two different conditions

The analysis will be done by calculating the sample mean difference for the two samples.

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

Statistics	Congruent	Incongruent
Count in each smaple	24	24
Sample Mean	14.051125	22.01591667
Median	14.3565	21.0175
Variance	12.66902907	23.01175704
Sample Standard Deviation	3.5594	4.7971

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.

The below graph shows the
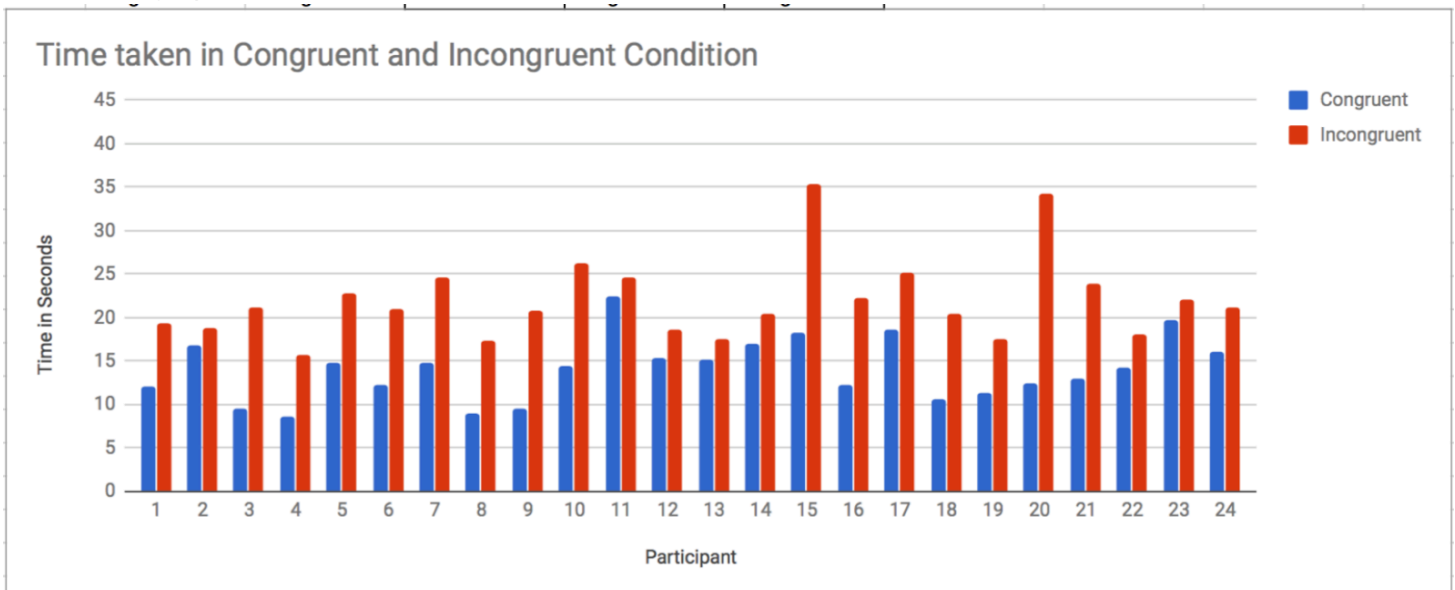
The graph shows that all the 24 participants took more time in reading the list of word in incongruent condition. From the graph, we can expect that irrespective of who takes the test, the time in the incongruent condition will always be more.

5. Now, perform the statistical test and report your results.

What is 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?

```
In [1]:  ## Will use python for calculations.
         ## importing libraries and reading data
         import pandas as pd
         import numpy as np

         data = pd.read_csv('stroopdata.csv')
         data
```

Out[1]:

	Congruent	Incongruent
0	12.079	19.278
1	16.791	18.741
2	9.564	21.214
3	8.630	15.687
4	14.669	22.803
5	12.238	20.878
6	14.692	24.572
7	8.987	17.394
8	9.401	20.762
9	14.480	26.282
10	22.328	24.524
11	15.298	18.644
12	15.073	17.510
13	16.929	20.330
14	18.200	35.255
15	12.130	22.158
16	18.495	25.139
17	10.639	20.429
18	11.344	17.425
19	12.369	34.288
20	12.944	23.894
21	14.233	17.960
22	19.710	22.058
23	16.004	21.157

The data is dependent samples as the same subject is put to two different conditions. Here we are going to conduct One tailed t-test. The first step will be to find the difference between the time taken by each person in two different condition.

```
In [2]:  ## difference between the time taken in each condition
         Time_diff = data['Incongruent'] - data['Congruent']
         Time_diff
```

```
Out[2]:  0         7.199
         1         1.950
         2        11.650
         3         7.057
         4         8.134
         5         8.640
         6         9.880
         7         8.407
         8        11.361
         9        11.802
         10        2.196
         11        3.346
         12        2.437
         13        3.401
         14       17.055
         15       10.028
         16        6.644
         17        9.790
         18        6.081
         19       21.919
         20       10.950
         21        3.727
         22        2.348
         23        5.153
         dtype: float64
```

Step 2- Calculate mean difference for the given two sample.
$$\bar{x_d} = \bar{x_i} - \bar{x_c}$$

where

$\bar{x_d}$ -> Mean difference between the two sample

$\bar{x_i}$ -> Sample mean for incongruent condition

$\bar{x_c}$ -> Sample mean for congruent condition

```
In [3]:  ## Mean difference between the two samples calculated using python

         mean_c = data['Congruent'].mean()
         mean_i = data['Incongruent'].mean()
         mean_d = mean_i - mean_c
         mean_d
```

```
Out[3]:  7.964791666666665
```

OR

```
In [4]:  Time_diff.mean()
```

```
Out[4]:  7.964791666666667
```

The second step will be to calculate sample standard deviation.

$$\sigma = \sqrt{\frac{\sum(x_d - \bar{x_d})^2)}{n - 1}}$$

where
σ -> Sample standard deviation
x_d -> Difference between the time taken by each participant for two conditions
$\bar{x_d}$ -> Mean of time difference between the two sample
n -> Number of participants in one sample

```
In [5]:  σ = Time_diff.std()
         σ
```

```
Out[5]:  4.864826910359056
```

The number of samples is 24 for each condition The next step is to calculate standard error. The formula for which is

$$SE = \frac{\sigma}{\sqrt{n}}$$

```
In [6]:  ## Calculating sample error using python code
         n = 24
         SE = σ/np.sqrt(n)
         SE
```

```
Out[6]:  0.99302863477834058
```

The t-districution will be drawn considering following factors-
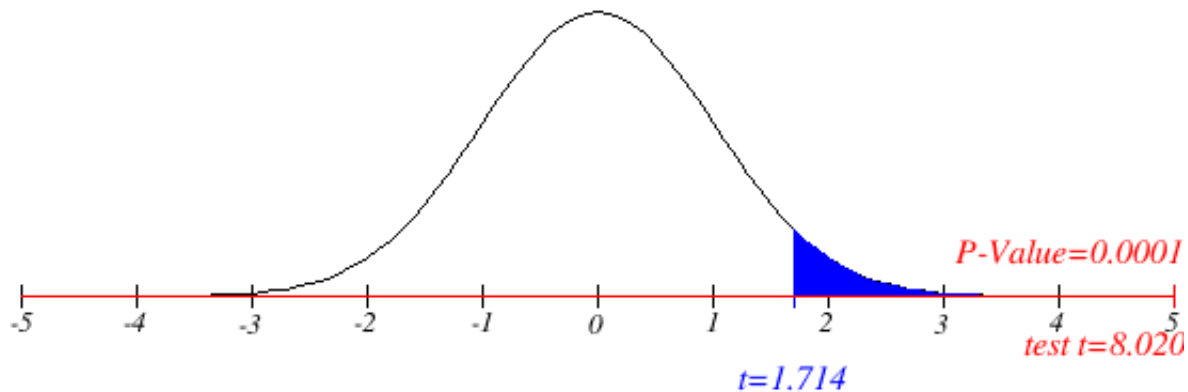
```
In [7]:  ##degrees of freedom
         D = 23

         ## alpha
         α = 0.05

         ## t critical at alpha 0.05
         tc = 1.714
```


The next step will be to find the t statistics for the sample mean difference (μd) -7.964791666666665. The formula fpr which is -

$$t = \frac{\bar{x_d} - 0}{\frac{\sigma}{\sqrt{n}}}$$

```
In [8]:  ## Calculating t using python
         t = (mean_d - 0)/SE
         t
```

Out[8]: 8.0207069441099552

Reject or Accept Null

As the t statistic is significantly higher than t-critical with $p < \alpha$, hence we can reject Null Hypothesis.

95% Confidence Interval

Here the confidence interval is for population mean difference, hence the point estimation is sample mean difference ($\bar{x_d}$) which is 7.965 For 95% confidence interval, t critical is (t_c) is $\pm$ 2.069

The formula is:

$$\bar{x_d} \pm (t_c * \frac{\sigma}{\sqrt{n}})$$

where
$\bar{x_d}$ -> Mean difference between the two sample
σ -> Sample standard deviation
n -> Number of partocipants in each sample
t_c -> t critical when $\alpha = 0.05$

```
In [9]:  upper_limit = mean_d + (2.069 * SE)
         lower_limit = mean_d - (2.069 * SE)
         print(upper_limit)
         print(lower_limit)

         10.019367912
         5.91021542131
```

95% CI for population mean difference is (5.91021542131,10.019367912).

Conclusion

From the sampling distribution it was expected that the time taken in the incongruemt condition will be more than the congruent condition. The statistical analysis shows that the t statistic for the sample mean difference between incongruent and congruent condition is significantly higher than t-critical with p < α and the populaton mean difference between time taken in the two condition ranges from 5.91 seconds to 10.02 seconds. As the range of time difference is positive, the analysis also shows that time taken by the population in incongruent condition will be more than that of congruent condition

6. Optional: What do you think is responsible for the effects observed? Can you think of an alternative or similar task that would result in a similar effect? Some research about the problem will be helpful for thinking about these two questions!

The result shows that the average time taken to read the colors in congruent condition is less than that in the incongruent condition. Which was expected, as the human brain automatically understands the meaning of words as a result of habitual reading.

A similar task could be the colours written in a foreign language are expected to be read faster than the colours written in English that conflict with the colour ink they are written in. This is because, we are habitual in reading English (or the language we know) and hence tend to read the words than to focus on color. However, when the same colored words are written in foreign laguage, our focus will be on what we know which will be color in this case.

Reference -
https://en.wikipedia.org/wiki/Stroop_effect (https://en.wikipedia.org/wiki/Stroop_effect)
www.amt.edu.au/statscomp/resources/2011/ACT/Senior_1.pptx