# Test a Perceptual Phenomenon

February 27, 2019

### 0.0.1 Analyzing the Stroop Effect

(1) What is the independent variable? What is the dependent variable?

**In this study , the independent variables are the congruent and incongruent words conditions while the dependent variable is the time it takes to name the ink color of the word presented by seonds.**

(2) What is an appropriate set of hypotheses for this task? Specify your null and alternative hypotheses, and clearly define any notation used. Justify your choices.

**Mathematically :**
**H0: Tcongruent >= Tincongruent ( This is the null hypothesis ; it's the condition that it's believed to be true before collecting data and going through the statistical analysis).**
**H1: Tocongruent < Tincongruent ( This is the alternative hypothesis ; it's the condition that would be proved to be true after collecting data and going through the statistical analysis).**
**Tcongruent : the time for a group given a congruent words conidtion to name the ink color of the word. Tincongruent : the time for a group given an incongruent words condition to name the ink color of the word. : the population mean for the time to name the ink color of the word.**
**Verbally :**
**The null hypothesis is that the time for a group given a congruent words condition,which means that the color of each word is the same as the meaning of the word, to name the ink colors is more than or equal to the time when they are given an incongruent words condition to name the ink color of the word.**
**The alternative hypothesis is that the time for a group given an congruent words condition,which means that the color of each word is same as the meaning of the word, to name the ink colors is less than the time when they are given an incongruent words condition to name the ink color of the word.**
**For this set of data and kind of study , I will perform the Dependent Paired Sample T Test because I am studying the effect of congruent and incongruent words conditions on one group of poeple to see the significant difference for the time to name the ink word color.**
**Paired Sample T-Test Assumptions: The dependent variable (time) is continuous. The data set is approximately normally distributed. Testing for normality is conducted on the differences between the two congruent and incongruent conditions, not the raw values of each condition itself.**

(3) Report some descriptive statistics regarding this dataset. Include at least one measure of central tendency and at least one measure of variability. The name of the data file is 'stroop-data.csv'.

```
In [202]: import pandas as pd
          # openning the stroopdata.csv file and store it in df
          df = pd.read_csv('stroopdata.csv')
          # reading the first five rows of the data set
          df.head()
```

```
Out[202]:     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
```

```
In [203]: # finding the number of rows in the data set
          df.shape[0]
```

```
Out[203]: 24
```

I can observe from the number of rows that the sample size for this study is very small.

```
In [204]: # checking if there are rows with missing values ( False means no rows with missing va
          df.isnull().sum().any()
```

```
Out[204]: False
```

```
In [205]: # checking if there are rows with zero values (zero means there are no rows with zero
          sum(df['Congruent']==0)+sum(df['Incongruent']==0)
```

```
Out[205]: 0
```

## 0.1 Measures of Central Tendency:

```
In [206]: df.describe()
```

```
Out[206]:          Congruent  Incongruent
          count   24.000000    24.000000
          mean    14.051125    22.015917
          std      3.559358     4.797057
          min      8.630000    15.687000
          25%     11.895250    18.716750
          50%     14.356500    21.017500
          75%     16.200750    24.051500
          max     22.328000    35.255000
```

I can observe that the sample size for this data set is very small , 24 rows. The mean time to name the ink word color for the group given congruent words condition is less than the time when given incongruent words condition (14.05 < 22.02). In addtion, I can observe from the min and max times for the group to name the ink word color given congruent words condition are less than when they are given incongruent words condition.

```
In [207]:   # finding the average time to name the ink color of a word for the group given congrue
            m1=df['Congruent'].mean()
            m1
```

```
Out[207]:   14.051124999999999
```

```
In [208]:   # finding the average time to name the ink color of a word for the same group given in
            m2=df['Incongruent'].mean()
            m2
```

```
Out[208]:   22.015916666666666
```

```
In [209]:   # finding the mean difference of time for the group in congruent and incongruent words
            m1-m2
```

```
Out[209]:   -7.9647916666666667
```

I can observe from the difference of the average time for the group given congruent words condition to name the ink colors and when they are given incongruent words condition to name the ink colors is -7.96 , which means that I have sufficient evidence to reject the null hypothesis [H0:  Tcongruent >=  Tincongruent ] in favor of the alternative hypothesis.The alternative hypothesis is that the time for a group given an congruent words condition,which means that the color of each word is same as the meaning of the word, to name the ink colors is less than the time when they are given an incongruent words condition to name the ink color of the word. However, this conclusion needs to be also imporved using the paried statistical test.

```
In [210]:   # finding the median time to name the ink colors for the group given congruent words c
            df['Congruent'].median()
```

```
Out[210]:   14.3565
```

```
In [211]:   # finding the median time to name the ink colors for the same group given incongruent
            df['Incongruent'].median ()
```

```
Out[211]:   21.0175
```

I can observe from the calculated mean and median that they are nealry the same for the group in congruent and incongruent words conditions.

```
In [212]:   # finding the mode time to name the ink colors for the group given congruent words con
            df['Congruent'].mode()
```

```
Out[212]:   0        8.630
            1        8.987
            2        9.401
            3        9.564
            4       10.639
            5       11.344
            6       12.079
            7       12.130
```

3

```
8      12.238
9      12.369
10     12.944
11     14.233
12     14.480
13     14.669
14     14.692
15     15.073
16     15.298
17     16.004
18     16.791
19     16.929
20     18.200
21     18.495
22     19.710
23     22.328
dtype: float64
```

**I can observe that the mode time to name the ink colors for the group given congruent words condition is 12 seconds.**

```
In [213]:   # finding the mode time to name the ink colors for the same group given incongruent wc
            df['Incongruent'].mode()

Out[213]:   0      15.687
            1      17.394
            2      17.425
            3      17.510
            4      17.960
            5      18.644
            6      18.741
            7      19.278
            8      20.330
            9      20.429
            10     20.762
            11     20.878
            12     21.157
            13     21.214
            14     22.058
            15     22.158
            16     22.803
            17     23.894
            18     24.524
            19     24.572
            20     25.139
            21     26.282
            22     34.288
            23     35.255
            dtype: float64
```
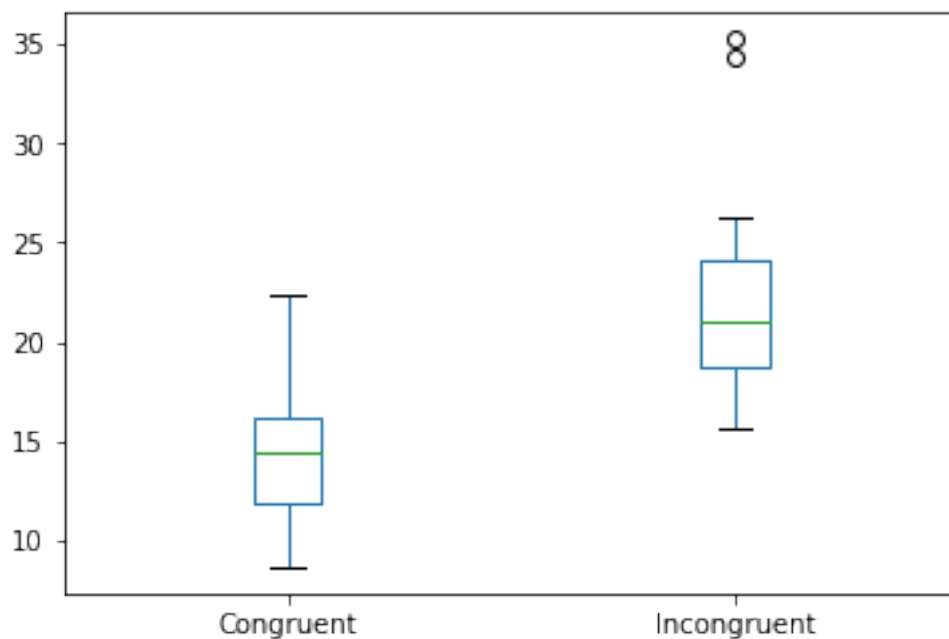
I can observe that the mode time to name the ink colors for the same group given incongruent words condition is the two values of 17 and 20 seconds.

As a final conclusion , since the mean , median and mode are almost the same under each words condition , I can say that the data set is almost noramlly distributed.

## 0.2 Measures of Variability :

```
In [214]: import matplotlib.pyplot as plt
          %matplotlib inline
          # plotting two box plots for congruent and incongruent words condition.
          df[['Congruent', 'Incongruent']].plot(kind='box')
```

```
Out[214]: <matplotlib.axes._subplots.AxesSubplot at 0x7f8656d46b70>
```



I can observe from the box plot above that there are two outliers for the group when they are given incongruent words condition. Since , the group study size is small , I will just keep both of them.

```
In [215]: # finding the range time to name the ink colors for the group given congruent words co
          df['Congruent'].max()- df['Congruent'].min()
```

```
Out[215]: 13.698000000000002
```

```
In [216]: # finding the range time to name the ink colors for the same group given incongruent u
          df['Incongruent'].max()- df['Incongruent'].min()
```

```
Out[216]: 19.568000000000001
```

5

I can observe that the time range to name the ink colors for the group given congruent words condition is less than the time range when they are given incongruent words condition.

```
In [217]:  # finding the standard deviation of time naming the ink word color for the group given
           df['Congruent'].std()

Out[217]:  3.5593579576451955

In [218]:  # finding the standard deviation of time naming the ink word color for the same group
           df['Incongruent'].std()

Out[218]:  4.7970571224691376
```

I can observe that how much time to name the color ink on average for each group member is different by 3.56 seconds from the the mean time of 14.05 seconds to name the ink word color for the whole group given congruent words condition.

I can also can observe that how much time to name the color ink on average for each group member is different by 4.80 seconds from the mean time of 22.02 seconds to name the ink word color for the whole group given incongruent words condition.

As a final conclusion , I can observe that the standard deviation of the time naming the ink word color for the group given congruent words condition is less than the standard deviation of time naming the ink word color when they are given incongruent words condition.
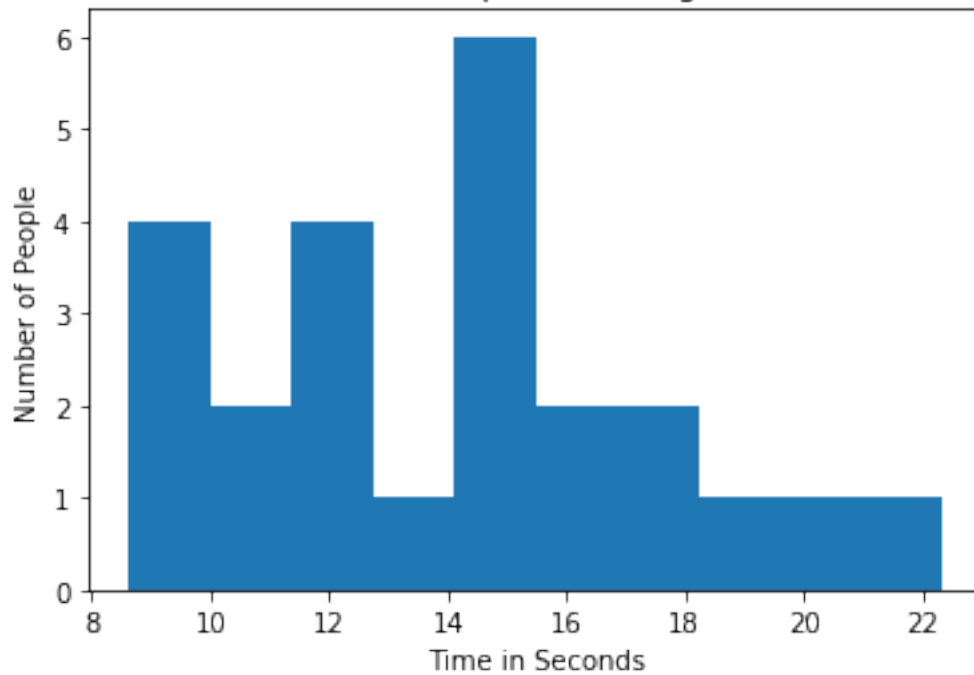
(4) Provide one or two visualizations that show the distribution of the sample data. Write one or two sentences noting what you observe about the plot or plots.

```
In [219]:  import matplotlib.pyplot as plt
           %matplotlib inline

           # plotting the time it take to name the ink word color for group given congruent words
           plt.hist(df['Congruent'])
           plt.title('Time Distribuation for a Group Given Congruent Words Condition')
           plt.xlabel('Time in Seconds')
           plt.ylabel('Number of People')

Out[219]:  Text(0,0.5,'Number of People')
```
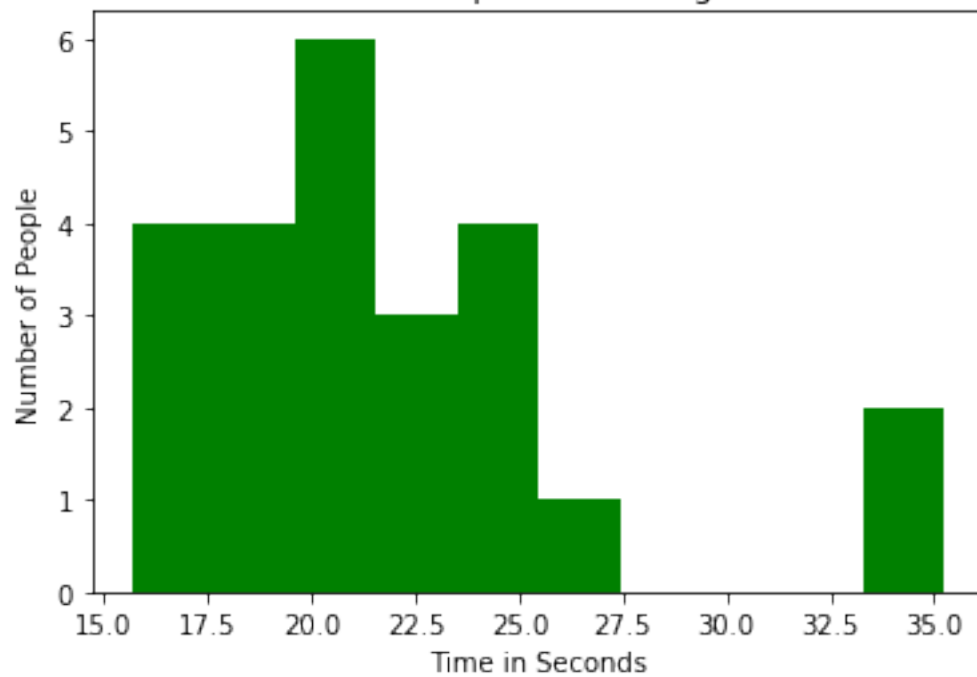
## Time Distribuation for a Group Given Congruent Words Condition



In [220]: *# plotting the time it take to name the ink word color for the same group given incong*
```
plt.hist(df['Incongruent'],facecolor='g')
plt.title('Time Distribuation for a Group Given Incongruent Words Condition')
plt.xlabel('Time in Seconds')
plt.ylabel('Number of People')
```

Out[220]: Text(0,0.5,'Number of People')

## Time Distribuation for a Group Given Incongruent Words Condition

```
In [221]: # plotting the time it take to name the ink word color for group given congruent words
          plt.hist(df['Congruent'])
          plt.hist(df['Incongruent'])
          plt.title('Time Distribuation for a Group Given Congruent and Incongruent Words Condit
          plt.xlabel('Time in Seconds')
          plt.ylabel('Number of People')
          labels= ["low","medium", "high"]
          plt.legend(['Congurent','Incongruent'])

Out[221]: <matplotlib.legend.Legend at 0x7f8656b7db00>
```

## Time Distribution for a Group Given Congruent and Incongruent Words Condition
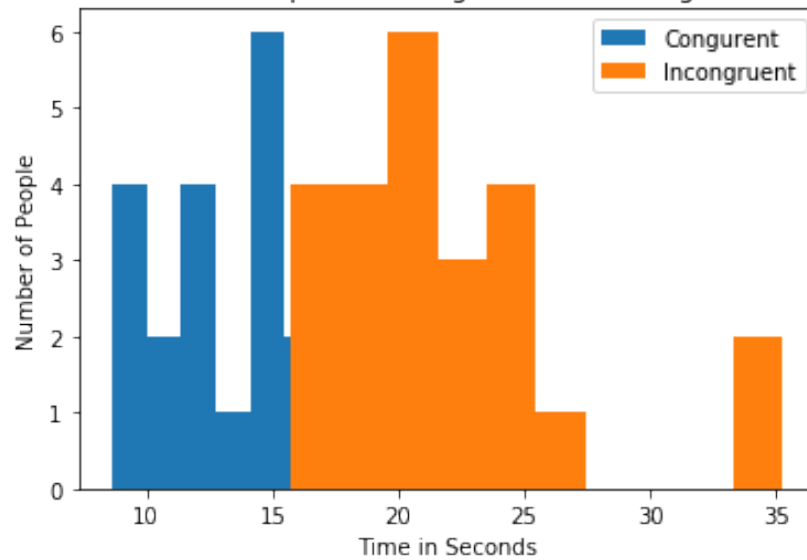


    **I can observe from the two histograms above that they are nearly normally distributed as their shape is so close to noraml distribuation than any other distribuation shape. I think the small size of the sample is affecting its shape somehow.**

(5) Now, perform the statistical test and report your results. What is your confidence level or Type I error associated with your test? What is your conclusion regarding the hypotheses you set up? Did the results match up with your expectations? **Hint:** Think about what is being measured on each individual, and what statistic best captures how an individual reacts in each environment.

    **To perform the Paired Sample T-Test , I will need to go into the following steps:**

```
In [222]: import numpy as np
          from scipy import stats

          # finding the sample size for the data set
          N = df.shape[0]
          N

Out[222]: 24

In [223]: # finding the sample standard deviation of the differences between the means
          sample_diff = df['Congruent']-df['Incongruent']
          sample_std = sample_diff.std()
          sample_std

Out[223]: 4.8648269103590556
```

```
In [224]: ## Calculating the t-statistics
          import math
          cong_mean = df['Congruent'].mean()
          incong_mean = df['Incongruent'].mean()
          t = (cong_mean - incong_mean)/(sample_std/math.sqrt(N))
          t
```

Out[224]: -8.020706944109957

```
In [225]: # finding the T critical value
          from scipy.stats import t
          # confidence level
          p = 0.95
          # degree of freedom
          deg = 23
          critical_value = t.ppf(p, deg)
          critical_value
```

Out[225]: 1.7138715277470473

```
In [226]: # performing the paried sample T test through the ready function to find also the p-va
          stats.ttest_rel(df['Congruent'], df['Incongruent'])
```

Out[226]: Ttest_relResult(statistic=-8.020706944109957, pvalue=4.1030005857111781e-08)

I can observe that when comparing the absolute value of the test statistic [-8.02] to the critical value of [1.71] that the test statistic is more than the critical value. Therefore, I have enough evidence to reject the null hypothesis in favor of the alternative hypothesis.

In addition, I can observe that the p-value is equal to zero which means it's less than the allowed type 1 error , 0.05. Therefore, I have enough evidence to reject the null hypothesis in favor of the alternative hypothesis; that states that the time for a group given an congruent words condition,which means that the color of each word is same as the meaning of the word, to name the ink colors is less than the time when they are given an incongruent words condition to name the ink color of the word.

(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!

After some reading on Wikipedia among several theories , I found that the ink word color and the word itself affects the speed of naming the ink word color. The human brain reads words faster than recognizing the ink word color. As a result, if an individual was given a word with a mismatched ink color , he or she will take more time to recognize the ink word color. On the other hand, if or she was given a word with a matched ink color , the speed of reconginzing the ink word color will be much fatser.

Another task similar to the stroop effect is the warped words. When words are printed in a way that is more difficult to read, the individual will take more time to recognize the word itself. The theory here that the way in which the word is printed affects the brains's reaction and the processing time to complete the given task.

## 0.3 Refrences

Link 1
 Link 2
 Link 3
 Link 4
 Link 5
 Link 6
 Link 7
 Link 8
 Link 9
 Link 10
 Link 11
 Link 12
 Link 13
 Link 14
 Link 15
 Link 16
 Link 17
 Link 18
 Link 19
 Link 20
 Link 21
 Link 22