# Test a Perceptual Phenomenon

August 31, 2018

## 1 Introduction

First we need to explore our data by printing a few examples and type of each variable. Then we

### 1.0.1 Analyzing the Stroop Effect

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

```
In [14]: %matplotlib inline
         import pandas as pd
         import matplotlib.pyplot as plt
         import plotly.plotly as py
         import seaborn as sns
         import scipy.stats as stats
```

The independent variable : The type of test ( congruent / incogruent )
The dependent variable : The response time in second
The first step is importing the data, stroopdata.csv, our file is a csv and can be imported into the dataframe.

```
In [15]: stroop_data = pd.read_csv('stroopdata.csv')
```

Let's do some exporatory analysis and the kind of data we will be working with.

```
In [16]: stroop_data.head(3)
```

```
Out[16]:    Congruent  Incongruent
         0     12.079       19.278
         1     16.791       18.741
         2      9.564       21.214
```

–write answer here–

(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.

We want to to evaluate the effect of the performance on the participants see whether the differences among samples is a result of congruent or incongruent tests.

Hypotheses:

H0 : The null hypothesis, that the congruent and incongruent populations are not significantly different (C  I = 0)

HA : The alternative hypothesis, that the congruent and incongruent populations are significantly different (C  I  0)

In this case, the sample size is below 30 and each individual participant is measure at two time points. This means we will have to conduct a dependant t-test for paired samples.

(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 [17]: stroop_data.describe()
```

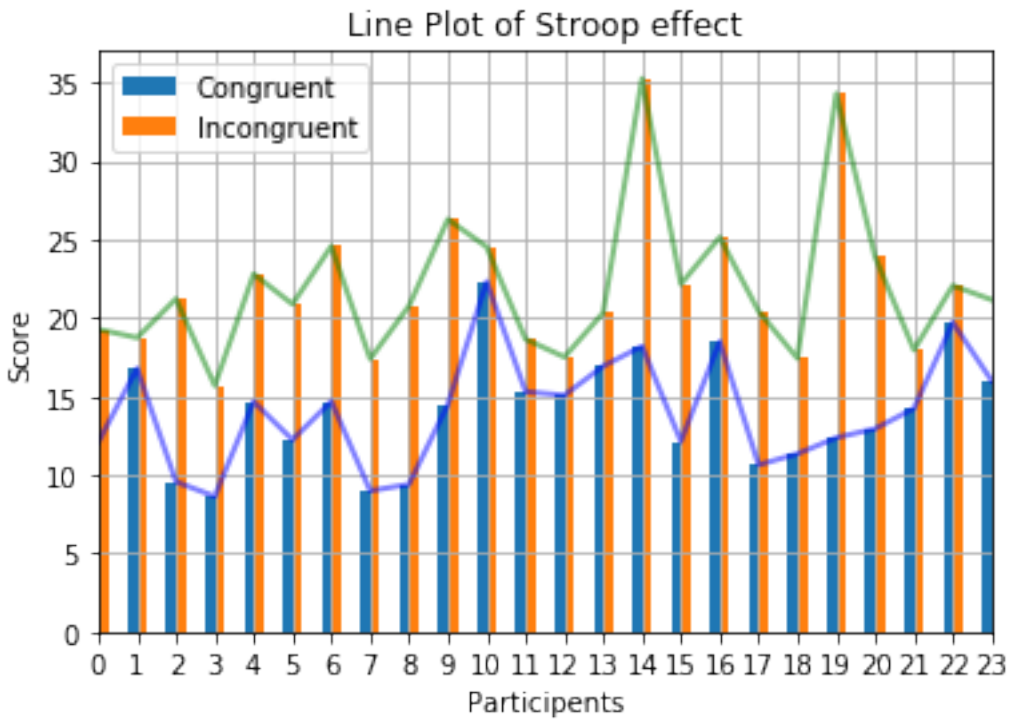
```
Out[17]:         Congruent  Incongruent
         count  24.000000    24.000000
         mean   14.051125    22.015917
         std     3.559358     4.797057
         min     8.630000    15.687000
         25%    11.895250    18.716750
         50%    14.356500    21.017500
         75%    16.200750    24.051500
         max    22.328000    35.255000
```

The congruent sample has a mean of xC = 14.051 seconds and a standard deviation C = 3.559 seconds. The incongruent sample has a mean of xI = 22.016 seconds and a standard deviation I = 4.797 seconds.

(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.

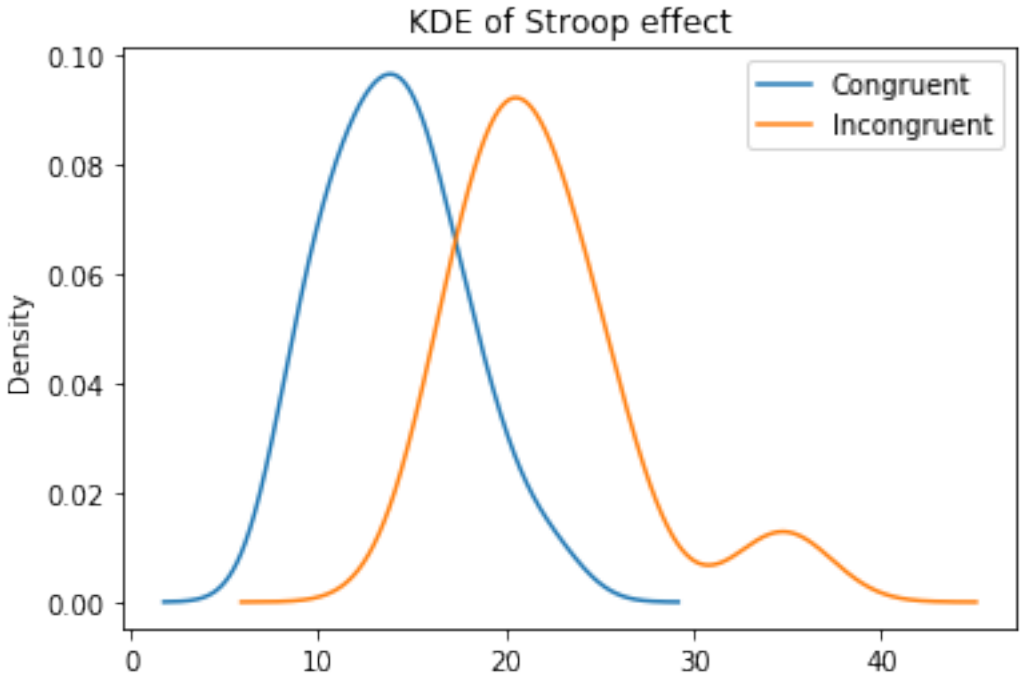
```
In [18]: stroop_data.plot(kind='bar')
         stroop_data.Congruent.plot(kind = 'line', color = 'b',label = 'Congruent',linewidth=2,a
         stroop_data.Incongruent.plot(color = 'g',label = 'Incongruent',linewidth=2, alpha = 0.5
         plt.xlabel('Participents')
         plt.ylabel('Score')
         plt.title('Line Plot of Stroop effect')
```

```
Out[18]: Text(0.5,1,'Line Plot of Stroop effect')
```




```
In [19]: stroop_data.plot.kde()
         plt.title('KDE of Stroop effect')

Out[19]: Text(0.5,1,'KDE of Stroop effect')
```



Both samples are normally distributed with a similar varianc. The incongruent task distribution of performance times seems to peak around 35 seconds.

(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.

```
In [21]: # Calculte the mean of each test
         cong_mean = stroop_data['Congruent'].mean()
         incong_mean = stroop_data['Incongruent'].mean()

         # Calculte a point estimate using (mean_congurent - mean_incongurant) / Average of diff
         estimate_point = cong_mean - incong_mean

         # Calculate Square of Sums of difference of sample means
         ssd = sum( ( (stroop_data['Congruent'] - stroop_data['Incongruent']) - estimate_point )

         # Calculate the variance of difference
         vd = ssd / (24 -1)

         # Calculate Standard Deviation of Differences
         sdd = vd ** (1/2)


         #  calculate the t-statistic for the difference
         tstatistic = estimate_point / ( sdd / ( 24 ** (1/2) ) )

         # Calculate the Degrees of Freedom (n -1)
         df = 23

         # Calculte the chances of seeing a result as extreme as the one we observed (known as t
         pvalue = stats.t.cdf(x = tstatistic, df = 23 ) * 2

         # Clacute t-critical for alpha level 0.005 (aiming to 99% confidence level)
         alpha = 0.005
         tcritical = stats.t.ppf(q=alpha, df=23)

In [22]: print("Degrees of Freedom : df =", df)
         print("Alpha- level : a =", alpha)
         print("t-critical values :  tc  =", -tcritical, tcritical)
         print("t-Statistic(df) : t(23) =", tstatistic)
         print("p-value : p =", pvalue)

Degrees of Freedom : df = 23
Alpha- level : a = 0.005
```

```
t-critical values :  tc  = 2.80733568377 -2.80733568377
t-Statistic(df) : t(23) = -8.02070694411
p-value : p = 4.10300058571e-08
```

The t-statistic is clearly smaller than the negative t-critical value and the critical reigion with a p-value < 0.0001.

**Therefore we can reject the Null Hypothesis with a 99% confidence interval and conclude there is a significant difference between congruent and incongruent tasks.**

These reuslts match what our observations and expectations that there is a statistically significant difference in the response.