Perceptual Phenomenon Test

January 22, 2019

0.0.1 Analyzing the Stroop Effect

(1) The independent variable and the dependent variable

- The independent variable is the Task condition: Congruent (the name of the color matches with the ink color), and Incongruent (the name of the color do not match with the ink color)
- The dependent variable is the Time (Time it takes to name the ink colors)

(2) The null and alternative hypotheses

An appropriate hypothesis for this task is:

- We want to approve that there is a difference between the in average reaction time of the incongruent condition and the average reaction time of the congruent condition.
- The null hypothesis (H0) assumes that there is no difference between the two averages, while the Alternative hypothesis (Ha) assumes that there is a difference between the two and that the average time of the task under the incongruent condition is longer than the average time of the task under the congruent condition

H0: Mc = Mi

Ha: Mc < Mi

- (Mc) is the average times of congruent task data and (Mi) is the average time of incongruent task data

To achieve this we can use a paired sample t-test (One Sample Paired t-test) since we are comparing the means of two groups that are dependent (Whether the mean difference between the two sets of observations is zero) and the same group is involved under two conditions

(3) Descriptive statistics regarding this dataset

```
In [1]: # import all necessary libraries and packages
        import pandas as pd
        import numpy as np
        import matplotlib.pyplot as plt
        import scipy.stats as stats
```

```
import seaborn as sns
sns.set_style('darkgrid')
% matplotlib inline

# Loading the data and print out a few lines
df = pd.read_csv('stroopdata.csv')
df.head()
```

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

In [2]:
```
# return useful descriptive statistics
df.describe()
```

Out[2]:

	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

In [3]:
```
# Calculate the mean for the Congruent data
con_mean = df['Congruent'].mean()
con_mean
```

Out[3]: 14.051124999999999

In [4]:
```
# Calculate the mean for the Incongruent data
incon_mean = df['Incongruent'].mean()
incon_mean
```

Out[4]: 22.015916666666666

In [5]:
```
# Calculate the variance for the Congruent data
df['Congruent'].var()
```

Out[5]: 12.669029070652176

In [6]:
```
# Calculate the variance for the Incongruent data
df['Incongruent'].var()
```

Out[6]: 23.011757036231884

In [7]:
```python
# Calculate the mean diffeernce between the two task conditions
mean_diff = incon_mean - con_mean
mean_diff
```

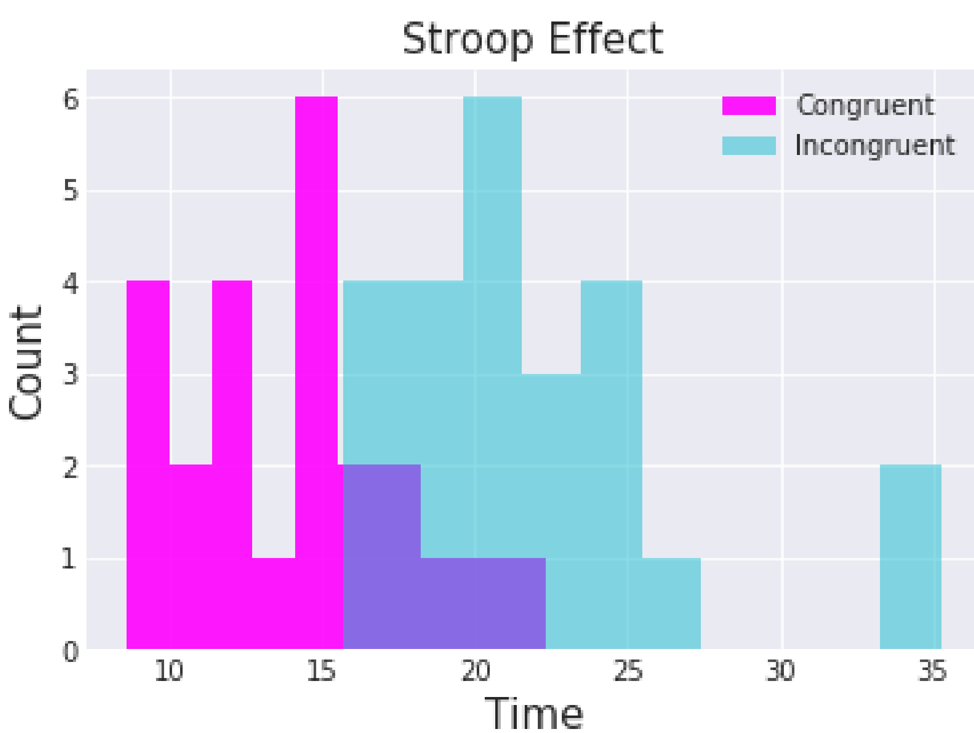
Out[7]: 7.9647916666666667

- As shown above, the average times of the 24 participants under the congruent condition and the incongruent condition is Xc=14.051125 and Xi=22.015917 respectively. The sample standard deviations of the two conditions are Sc=3.559358 and Si=4.797057. The difference between the two average times is Xi - Xc = 7.9647916666666667

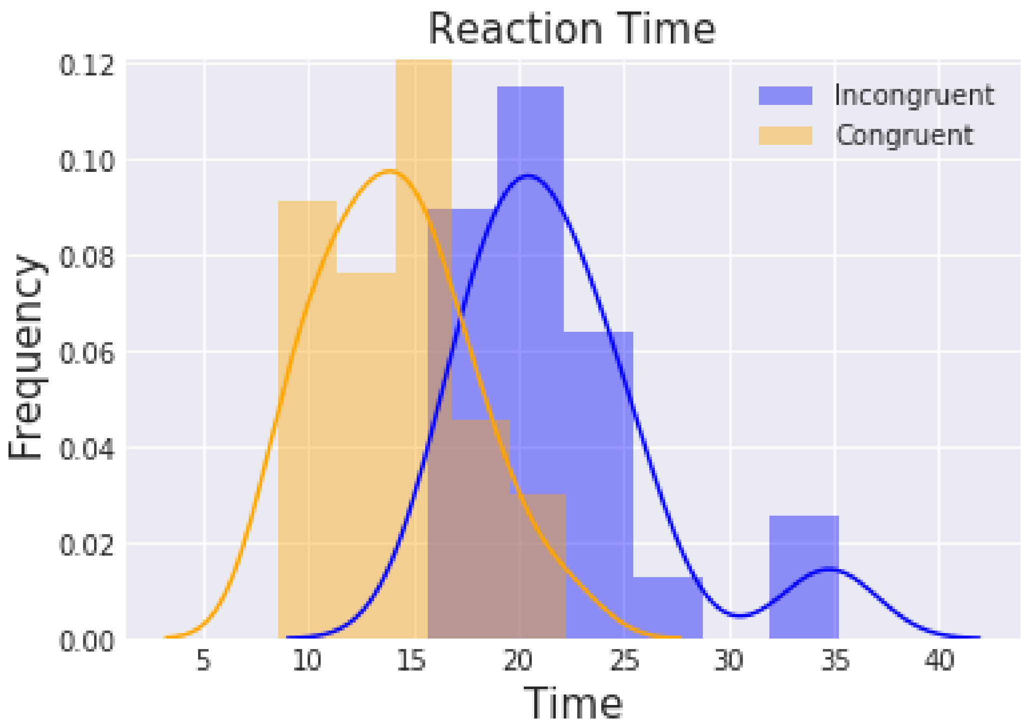
(4) Visualizations that show the distribution of the sample data

In [8]:
```python
# plot a histogram for the reaction times of the Congruent task and Incongruent task
plt.hist(df['Congruent'], histtype = 'bar',
         color = 'magenta', alpha=0.9, label = 'Congruent')
plt.hist(df['Incongruent'], histtype = 'bar',
         color = 'tab:cyan', alpha=0.5, label = 'Incongruent')

plt.xlabel('Time', fontsize = 15)
plt.ylabel('Count', fontsize = 15)
plt.title('Stroop Effect', fontsize = 15)
plt.legend()
plt.grid(True)
plt.show()
```



In [9]:
```
# Visualize the distribution of a dataset
sns.distplot(df['Incongruent'], color = "blue", label = "Incongruent")
sns.distplot(df['Congruent'], color = 'orange', label = "Congruent")
plt.xlabel("Time", fontsize = 15)
plt.ylabel("Frequency", fontsize = 15)
plt.title("Reaction Time", fontsize = 15)
plt.legend()
plt.show();
```

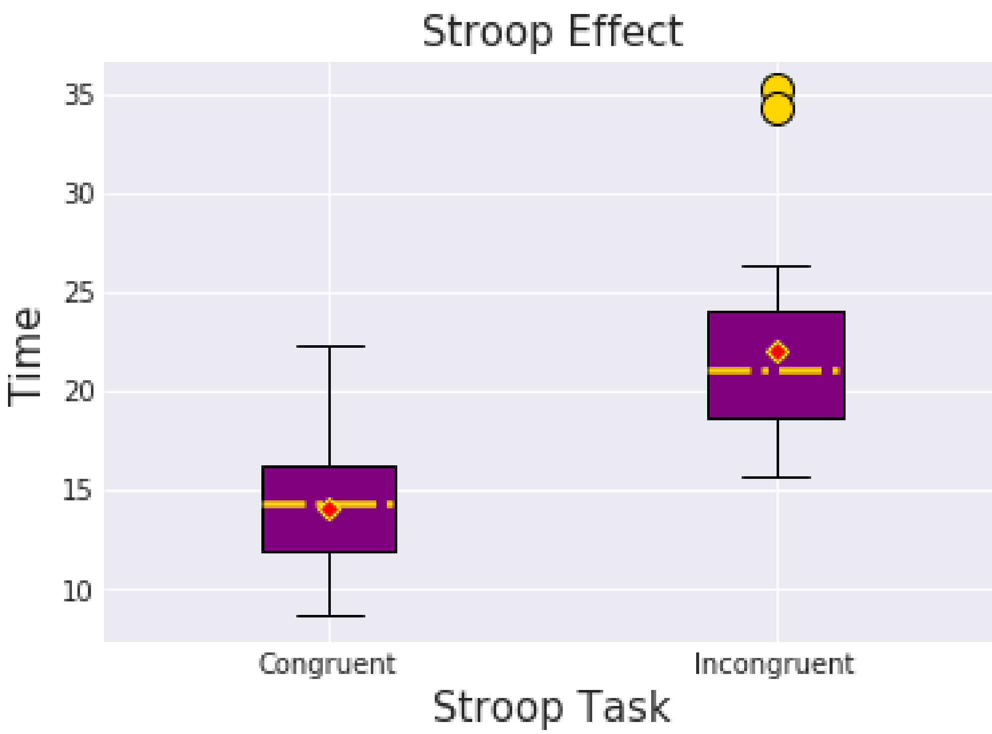


- The histograms show that the two groups have a significant difference in times, which is consistent with the statistics generated earlier. We can see that the mean is different for both the distributions

In [10]:
```
# Boxplot of Reaction Time of the Task under the congruent condition and the incongruen
flierprops = dict(marker='o', markerfacecolor='gold',
                  markersize=12, linestyle='none')
meanpointprops = dict(marker='D', markeredgecolor='yellow',
                      markerfacecolor='red')
medianprops = dict(linestyle='-.', linewidth=2.5, color='gold')
label = ['Congruent', 'Incongruent']
plt.boxplot([df['Congruent'], df['Incongruent']], widths = 0.3,
            patch_artist = True, labels = label, showmeans = True,
            flierprops = flierprops, meanprops = meanpointprops,
```

```
                 medianprops = medianprops,
                 boxprops=dict(facecolor='purple', color='black'))

plt.title('Stroop Effect')
plt.xlabel('Stroop Task', fontsize = 15)
plt.ylabel('Time', fontsize = 15)
plt.title('Stroop Effect', fontsize = 15)
plt.show()
```



- The boxplot shows that the reaction time under the Incongruent condition is longer than that of the Congruent condition, and the two tasks have different ranges and different median times - with the Incongruent task presents much longer times. Also, the distribution of Incongruent data is likely to be positively skewed due to the two outliers

(5) Statistical test

In [11]:
```
#find the diffenence of each data
df['difference'] = df['Congruent'] - df['Incongruent']
df.head()
```

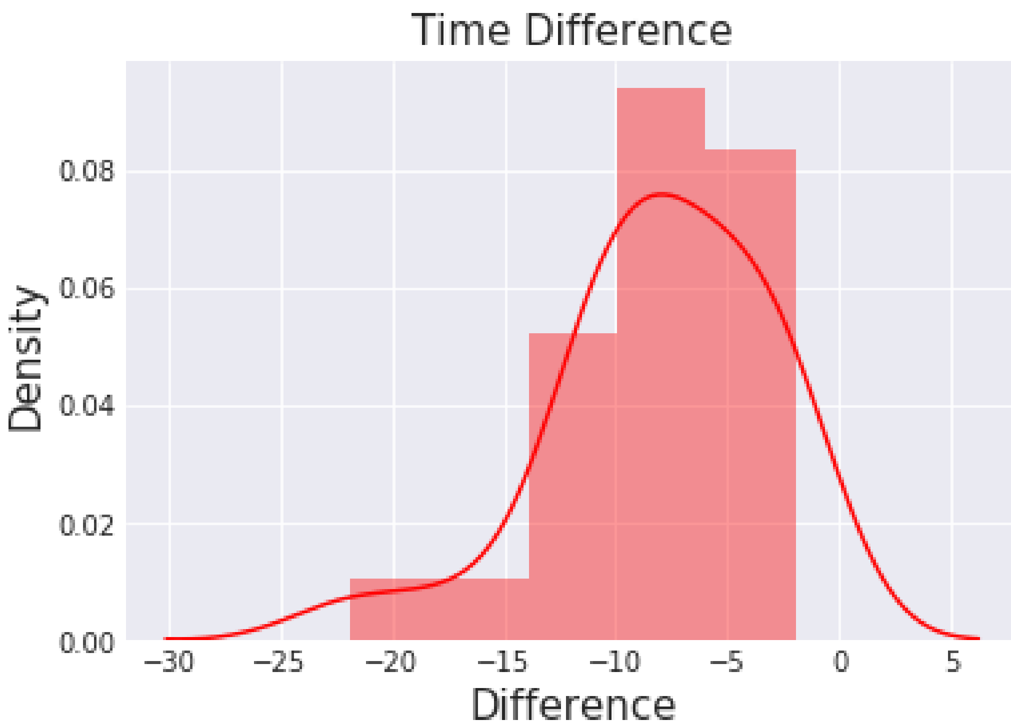
Out[11]:

	Congruent	Incongruent	difference
0	12.079	19.278	-7.199
1	16.791	18.741	-1.950
2	9.564	21.214	-11.650

```
3      8.630       15.687      -7.057
4     14.669       22.803      -8.134
```

```
In [12]: # plot the distribution of time difference
         sns.distplot(df['difference'], color = "red")

         plt.xlabel("Difference", fontsize = 15)
         plt.ylabel("Density", fontsize = 15)
         plt.title("Time Difference", fontsize = 15)
         plt.show();
```



- Based on the plot, most of the values (Differences) that fall under the curve are negative, which indicates that the reaction time of the task under the incongruent condition is longer than that of the congruent condition

```
In [13]: #find t-critical value for 95% confidance interval and
         # 23 degree of freedom for the paired t test
         t_critical = stats.t.ppf(0.95, 23)
         t_critical

Out[13]: 1.7138715277470473
```

```
In [14]: # Calculate the t-statistic and p-value
         t, p = stats.ttest_rel(df.Congruent, df.Incongruent)
         t, p

Out[14]: (-8.020706944109957, 4.1030005857111781e-08)
```

0.0.2 Conclusion

- With an level of 0.05, the t-statistics is equal to -8.020706944109957, the abs of t-statistics is less than the t-critical value (1.7138715277470473)
- The p-value is nearly zero (4.1030005857111781e-08), which is less than (0.05), and therefore based on both the t-statistic and the p-value of the test, we reject the null hypothesis (H0). Hence we can conclude that there is a difference between the reaction time where the time under the incongruent condition is significantly longer than that of the congruent condition #### The results match up with my expectation, the Stroop effect indeed interferes the reaction time associated with the task

(6) Final thoughts

- The human brain tends to read first before considering the word color, and naming the color of a word will need more effort thus more time
- An alternative or similar task would be naming the true color of fruits being displayed in different colors

0.0.3 References

- Stroop effect
- Paired t-Test
- matplotlib.pyplot.hist
- How to Code the Student's t-Test from Scratch in Python