

# Test a Perceptual Phenomenon

October 27, 2017

## 0.0.1 Analyzing the Stroop Effect

Perform the analysis in the space below. Remember to follow [the instructions](#) and review the [project rubric](#) before submitting. Once you've completed the analysis and write up, download this file as a PDF or HTML file and submit in the next section.

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

The independent variable is whether the word is congruent or not. The dependent variable is the time it takes for each participant to say the color of the ink.

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

I expect to form a 2 tailed t test because I want to see if the words being incongruent had a significant effect in the time it took to say the color of the ink. Since I don't know the standard deviation of the whole population of people and I only have a sample of people, I have to use a t-test and not a z-test. Furthermore, since my sample size is less than 30, it's better if I use a t-test and not a z test. I'm also using a dependent T-test because the Samples are related to each other, since it's the same participants only with the different condition of saying Incongruent words. The hypotheses would be:

H0: There is no significant difference in the population average amount of time it takes to state the colors of the words in a incongruent condition if compared with a congruent condition. (mean of incongruent words = mean of congruent words)

H1: There is a significant difference in the population average amount of time it takes to state the colors of the words in a incongruent condition if compared with a congruent condition. (mean of incongruent words != mean of congruent words)

- (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 [7]: import pandas as pd
import scipy.stats as stats
import matplotlib as plt
import numpy as np
data = pd.read_csv('stroopdata.csv')
congruent = data['Congruent'].values.tolist()
```

```

incongruent = data['Incongruent'].values.tolist()
mean = data.mean()
median = data.median()
std = data.std()
iqr_c = stats.iqr(congruent)
iqr_i = stats.iqr(incongruent)
print(mean)
print(median)
print(std)
print(iqr_c)
print(iqr_i)

```

```

Congruent      14.051125
Incongruent    22.015917
dtype: float64
Congruent      14.3565
Incongruent    21.0175
dtype: float64
Congruent      3.559358
Incongruent    4.797057
dtype: float64
4.3055
5.33475

```

The mean of Congruent words is 14.0511 s and Incongruent words is 22.0159 s. The median of Congruent words is 14.3565 s and Incongruent words is 21.0175 s. The standard deviation of Congruent words is 3.5594 and Incongruent words is 4.7971. The interquartile range of Congruent words is 4.3055. The interquartile range of Incongruent words is 5.3348.

- (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.

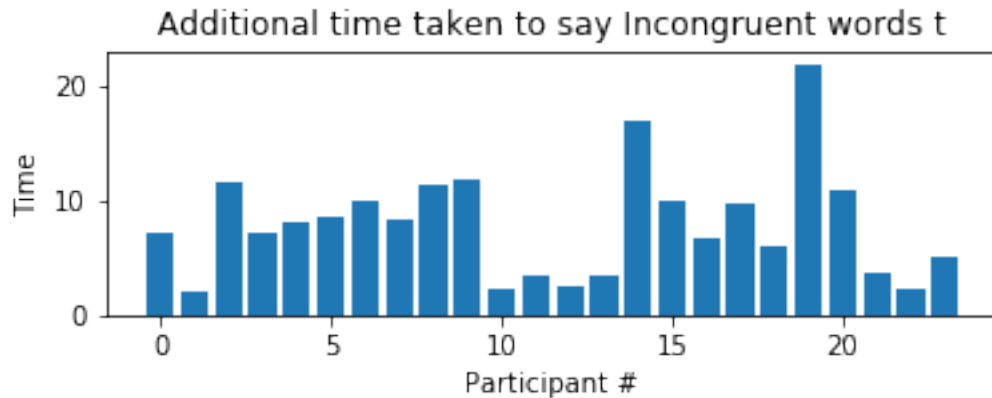
```
In [90]: import numpy as np
```

```

diff = data["Incongruent"].values - data["Congruent"].values
plt.subplot(211)
plt.bar(np.arange(24), diff)
plt.xlabel("Participant #")
plt.ylabel("Time")
plt.title("Additional time taken to say Incongruent words t")

```

```
Out[90]: Text(0.5,1,'Additional time taken to say Incongruent words t')
```



In the graph above, we see the Additional time taken by each Participant for saying Incongruent words. These numbers were calculated by doing Incongruent Time - Congruent Time. Obviously, all the numbers are positive, meaning everyone took longer saying Incongruent words.

- (5) Now, perform the statistical test and report the results. What is the 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]: import pandas as pd
import scipy.stats as stats
import matplotlib as plt
import numpy as np
data = pd.read_csv('stroopdata.csv')
congruent = data['Congruent'].values.tolist()
incongruent = data['Incongruent'].values.tolist()
stats.ttest_rel(congruent,incongruent)

Out[1]: Ttest_relResult(statistic=-8.020706944109957, pvalue=4.1030005857111781e-08)

In [17]: import pandas as pd
import scipy.stats as stats
import matplotlib as plt
import numpy as np
data = pd.read_csv('stroopdata.csv')
congruent = data['Congruent'].values.tolist()
incongruent = data['Incongruent'].values.tolist()
mean = data.mean()
median = data.median()
std = data.std()
smd = mean['Incongruent']-mean['Congruent']
tcrit = stats.t.ppf(q = 0.975, df=23)
sem = (((data['Incongruent'].values-data['Congruent'].values).std())/np.sqrt(24))
CI_lower = smd - tcrit * sem
CI_high = smd + tcrit * sem
print(CI_lower, CI_high)
```

5.95380734675 9.97577598658

```
In [7]: stats.t.ppf(q = 0.975, df=23)
```

```
Out[7]: 2.0686576104190406
```

Since the p-value,  $4.1030 \times 10^{-8}$  is less than an Alpha level of 0.05, we can reject the Null hypothesis. The T Critical Value is 2.0687, which is less than the T Statistic of 8.0207 (after inverting the sign in the python calculation). The Confidence Interval is (5.9538, 9.9758). In conclusion, it is safe to say that words being Incongruent had an effect (made people take longer) on the time it took to say the word. I was not surprised because it would make sense that when a word is Incongruent, the brain has to take more time to say the word compared with Congruent words.

I referred to Matplotlib library, Stack overflow, and pandas library websites for doing this project.