# Test a Perceptual Phenomenon

October 13, 2018

### 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, upload that PDF/HTML into the workspace here (click on the orange Jupyter icon in the upper left then Upload), then use the Submit Project button at the bottom of this page. This will create a zip file containing both this .ipynb doc and the PDF/HTML doc that will be submitted for your project.

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

--Independent Variable: Response Time; Dependent Variables: Congruent, Incongruent--

(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.

--Null: Ho: = 0
-- = Mean Population Congruent Time
-- 0 = Mean Population Incongruent Time
--Alternative: H1: 0 --
--These hypotheses were chosen so that the sample data provided could be used to make claims about the population's congruent and incongruent times as a whole. The following test will be two-tailed and be attempting to prove that response times for the two tests are not equal.

(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'.

```python
In [7]: import pandas as pd
        df = pd.read_csv('stroopdata.csv')
        df.describe()
```

```
Out[7]:        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
```

1

```
In [8]: df.median()

Out[8]: Congruent      14.3565
        Incongruent    21.0175
        dtype: float64

In [9]: df.var()

Out[9]: Congruent      12.669029
        Incongruent    23.011757
        dtype: float64

In [10]: df.skew()

Out[10]: Congruent      0.41690
         Incongruent    1.54759
         dtype: float64

In [11]: df.corr()

Out[11]:              Congruent  Incongruent
         Congruent      1.00000      0.35182
         Incongruent    0.35182      1.00000

In [12]: df.cov()

Out[12]:              Congruent  Incongruent
         Congruent    12.669029     6.007123
         Incongruent   6.007123    23.011757
```

### 0.0.2 Results Summary

In order to use descriptive statistics to analyze the csv file, I called upon some simple functions in python. I first used the describe() function, which demonstrates the difference in mean time for congruent and incongruent word lists in the stroop test. This shows that the mean time for incongruent word lists was about 8 seconds longer. This difference remains consistent when looking at the min and interquartile ranges for each. The main deviation is that the maximum time for incongruent grows to nearly 13 seconds longer than its counterpart. I then called upon the median() function which offered a similar result of seven seconds difference. These metrics to this point were depicting a certain statistical significance between the time it took to correctly analyze the incongruent word list and the congruent. I then called the var() function to check the spread of the data. Given that the variance is a high number, we can conclude that the data has a large spread, and is not clustered around the mean. The skew() function helps us determine whether the data is normally distributed (how close the skew result is to zero). The congruent data shows a more normal distribution, and the incongruent data shows a greater skew to the left of the mean. I followed that with the corr() function to analyze the correlation between the two variables. The results show a positive correlation of .35. This means for every increase of 1 in the time it took for the participant to read the congruent word list, there would be an increase of .35 in the time for the incongruent test. Finally, I used the cov() function to look at the level the two variables varied together. This shows similar results to our corr() function and confirms previous findings.

(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 [13]: import matplotlib.pyplot as plt

         labels = 'Congruent', 'Incongruent'
         sizes = df.Congruent.mean(), df.Incongruent.mean()
         colors = ['red', 'blue']
         explode = (0, 0.1)  # explode Incongruent slice
         plt.title('Stroop test')
         plt.pie(sizes, explode=explode, labels=labels, colors=colors, shadow=True, startangle=1

         plt.axis('equal')

         plt.show()

<matplotlib.figure.Figure at 0x7f205b333b70>
```

   --Shown above is the Stroop Test results of the mean Congruent and Incongruent test times. This pie chart most accurately reflects the large difference in times between the two tests, which proved consistent in our descriptive analysis --

(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 [14]: from scipy import stats
         stats.ttest_rel(df.Congruent, df.Incongruent)

Out[14]: Ttest_relResult(statistic=-8.020706944109957, pvalue=4.1030005857111781e-08)
```

   --Given that the observations were of a sample and not the general population, I conducted a t test instead of a z test. In this case I used a dependent sample t test due to the fact that the observations were of a sample and the variables were relatied. Due to the fact that there are 23 observations, this gives us 22 degrees of freedom. The T test shows us that there is a -8.02 difference between the Congruent and Incongruent test response times. While this is a significant difference, due to the small sample size, the difference can only be used as a jumping-off-point for further research. Given the size of the p value, we can expect that the Incongruent results will consistently be better than the Congruent results into the future. The probability of rejecting the null given the hypothesis is true. Given that the p value is less than 5%, we reject the null hypothesis.--
   --The assumptions using this test were that the dependent variable is continuous, normally distributed, and did not contain any outliers--

(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!

   --write answer here--