

Test a Perceptual Phenomenon

March 21, 2019

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?

--The independent variable: (Used to predict the response) - Condition i.e. Congruent or Incongruent--
--The dependent variable: (the variable we are interested in predicting) - Time --

(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.

--H0(Null Hypothesis) - Average time taken by the incongruent words is lesser or equal to the average time taken by congruent words--
--H1(Alternative Hypothesis) - Average time taken by the incongruent words is more than the average time taken by the congruent words--

--Please see the pdf for more details on notation--

(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 [71]: # Perform the analysis here
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
df = pd.read_csv("stroopdata.csv")
df.head(5)

#measure of central tendency - Median
median_c = df['Congruent'].median()

median_ic = df['Incongruent'].median()
```

```

#measure Mean
mean_c = df['Congruent'].mean()
mean_ic = df['Incongruent'].mean()

#measure of variability - variance
var_c = df['Congruent'].var()
var_ic = df['Incongruent'].var()

print('Median for Congruent Words:',median_c,'\nMedian for Incongruent Words: ',median_ic)
print('Mean for Congruent Words:',mean_c,'\nMean for Incongruent Words: ',mean_ic)
print('Variance for Congruent Words:',var_c,'\nVariance for Incongruent Words: ',var_ic)

```

```

Median for Congruent Words: 14.3565
Median for Incongruent Words: 21.0175
Mean for Congruent Words: 14.051125
Mean for Incongruent Words: 22.0159166667
Variance for Congruent Words: 12.6690290707
Variance for Incongruent Words: 23.0117570362

```

--The mean < median for congruent words the graph is left skewed. The mean > median for incongruent data the graph is right skewed--

- (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 [48]: # Build the visualizations here
import pandas as pd
from math import sqrt
import matplotlib.pyplot as plt
%matplotlib inline

#minimum of congruent
df['Congruent'].min()

Out[48]: 8.6300000000000008

In [49]: #maximum of congruent
df['Congruent'].max()

Out[49]: 22.328000000000003

In [50]: q3 = df['Congruent'].quantile(0.75)
q3

Out[50]: 16.20075

In [52]: q1 = df['Congruent'].quantile(0.25)
q1

```

```
Out[52]: 11.89525
```

```
In [53]: #IQR Range  
IQR = q3 - q1  
IQR
```

```
Out[53]: 4.3054999999999986
```

```
In [54]: q3in = df['Incongruent'].quantile(0.75)  
q3in
```

```
Out[54]: 24.0515
```

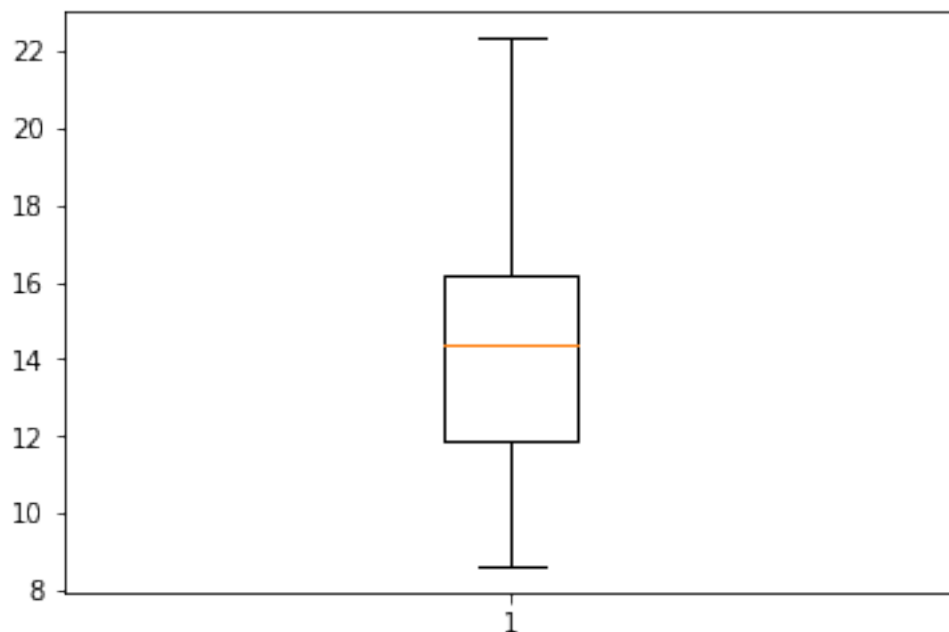
```
In [29]: q1in = df['Incongruent'].quantile(0.25)  
q1in
```

```
Out[29]: 18.71675
```

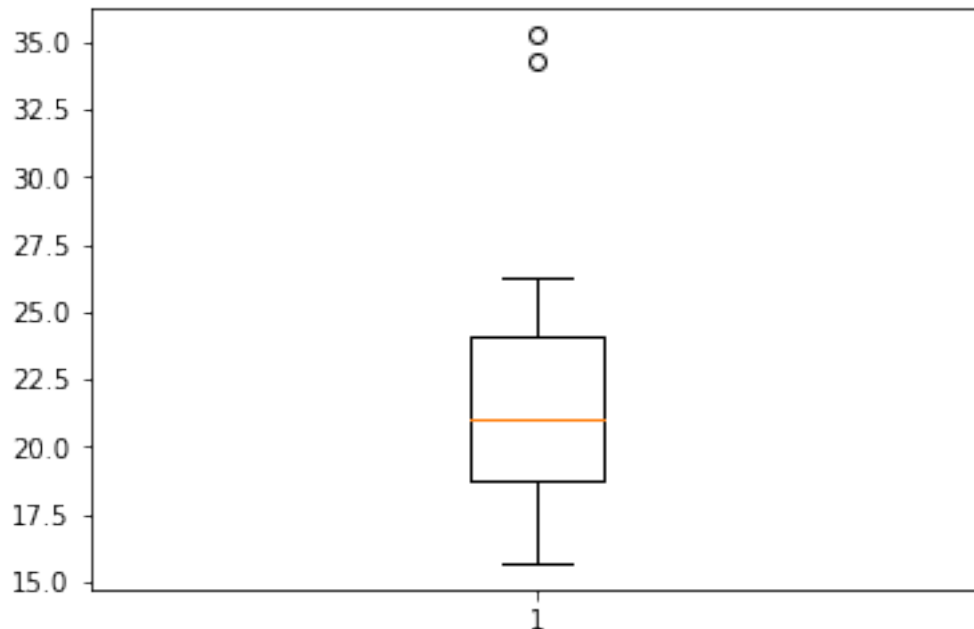
```
In [32]: IQRic = q3in - q1in  
IQRic
```

```
Out[32]: 5.33475
```

```
In [56]: plt.boxplot(df['Congruent'])  
plt.show()
```



```
In [33]: plt.boxplot(df['Incongruent'])  
plt.show;
```



--The congruent words data has no outliers but has a larger spread indicating the values lie away from the center of the data. The incongruent words data has two outliers but has smaller spread indicating the values lie closer to the center of the data.--

- (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.

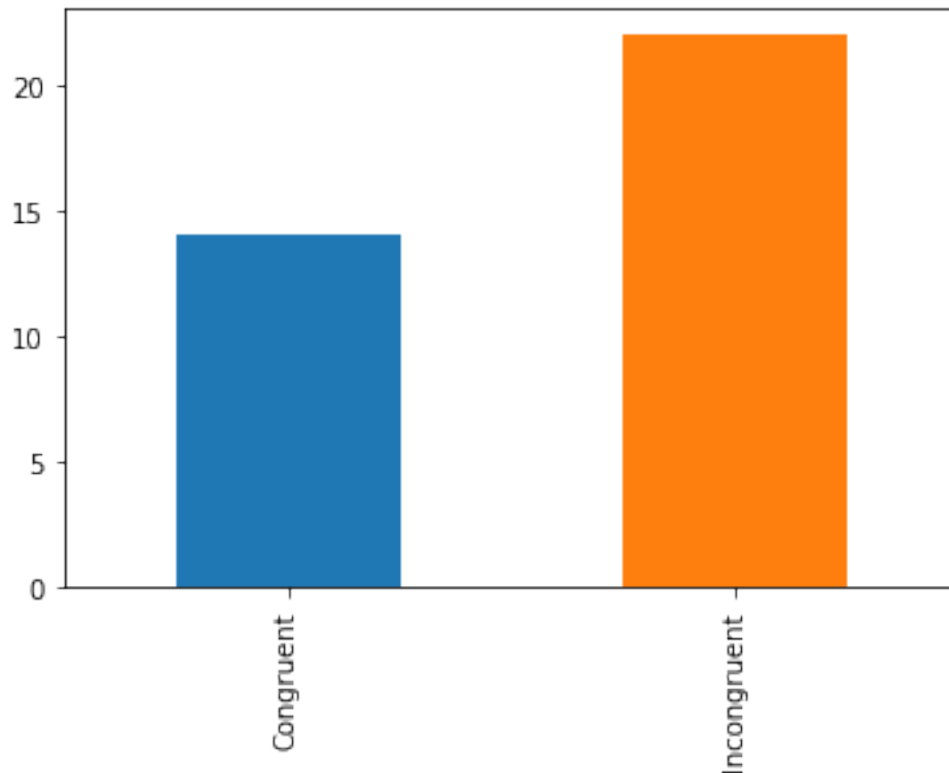
```
In [57]: # Perform the statistical test here
import scipy.stats as stats
df = pd.read_csv("stroopdata.csv")
df.head(5)
```

```
Out[57]:
```

	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 [37]: congruent = df['Congruent'].values.tolist()
incongruent = df['Incongruent'].values.tolist()
```

```
In [58]: df.mean().plot('bar');
```



```
In [59]: stats.ttest_rel(congruent,incongruent)
```

```
Out[59]: Ttest_relResult(statistic=-8.020706944109957, pvalue=4.1030005857111781e-08)
```

--p value is lesser than 0.05, the results are statistically significant and we can reject the null hypothesis--

--Type 1 Error: Average time taken by the incongruent words is more than the average time taken by congruent words when actually, average time taken by congruent words are more than or equal to the average time taken by the incongruent words. --

- (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--