Test a Perceptual Phenomenon

September 30, 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: The colour associated with the text.
Dependent Variable: Response time to read the text.

(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 Hypothese: (Time to read congruent words) - (Time to read an incongruent words) = 0
Alternative Hypothese: (Time to read an incongruent words) - (Time to read congruent words) 0

(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 [16]: import pandas as pd
         df = pd.read_csv('stroopdata.csv')
         df_differ = df['Incongruent'] - df['Congruent']
         df_diff_avg = df_differ.mean()
         df_diff_median = df_differ.median()
         #Median and Mean, both are measures of central tendency
         df_diff_avg, df_diff_median
```

```
Out[16]: (7.9647916666666658, 7.666499999999999)
```

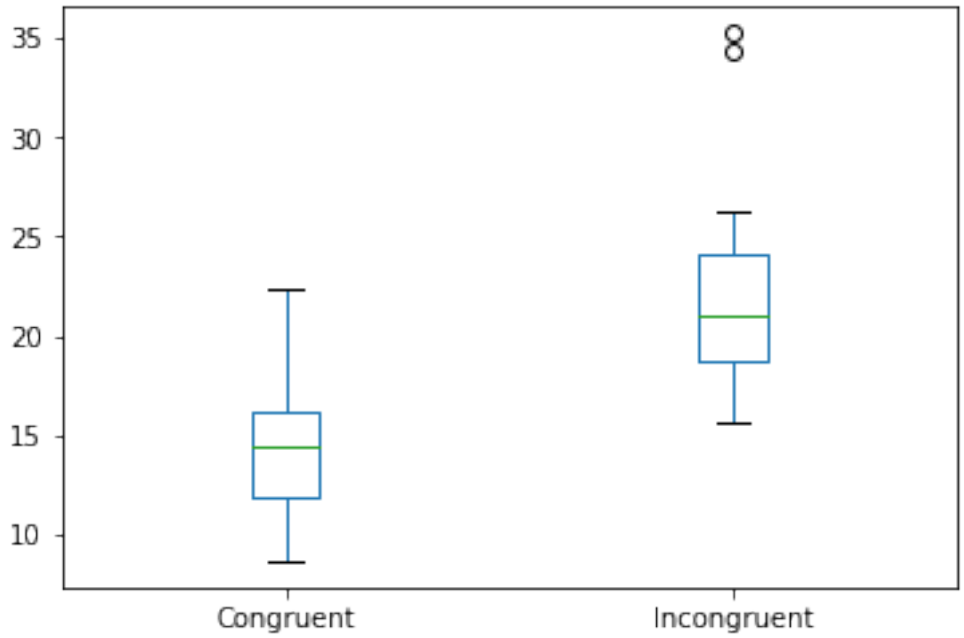
```
In [17]: diff_range = df['Incongruent'].max() - df['Congruent'].max()
         diff_std = df['Incongruent'].std() - df['Congruent'].std()
         #Range and Standard Deviation, both are measures of variability
         diff_range, diff_std
```

```
Out[17]: (12.927, 1.2376991648239422)
```

(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]: # Build the visualizations here
         import matplotlib.pyplot as plt
         df.plot(kind='box')
```

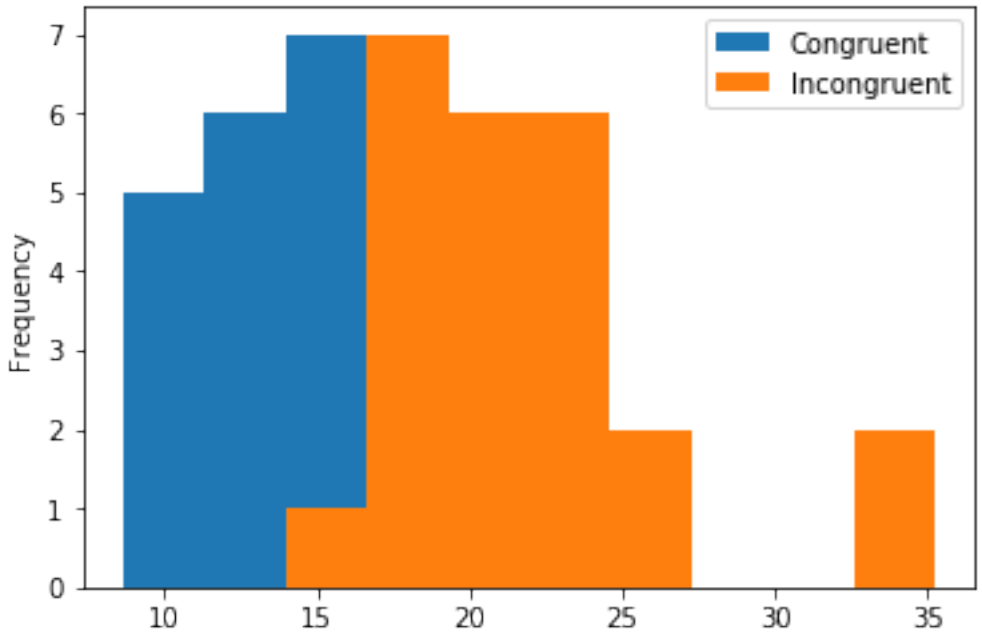
```
Out[18]: <matplotlib.axes._subplots.AxesSubplot at 0x7f60f63cf390>
```



From looking at the below box plot, one can see that median completion time of Incongruent group is higher. The box plot shows that, congruent study has higher range of variability as compared to incongruent.

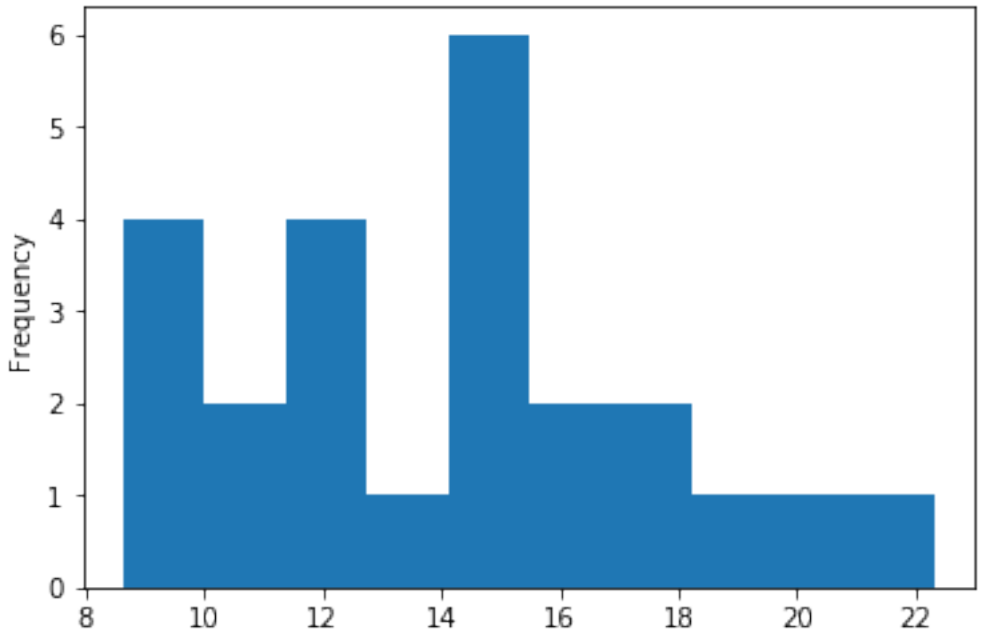
```
In [19]: df.plot(kind="hist")
```

```
Out[19]: <matplotlib.axes._subplots.AxesSubplot at 0x7f60f63c49e8>
```



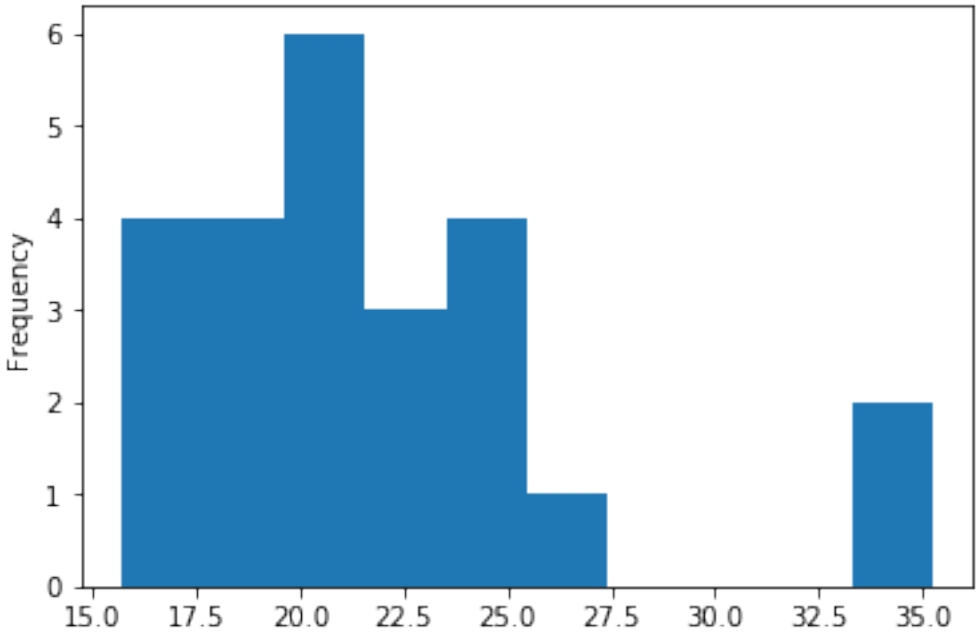

```
In [20]: df['Congruent'].plot(kind='hist')
```

```
Out[20]: <matplotlib.axes._subplots.AxesSubplot at 0x7f60f63603c8>
```



```
In [21]: df['Incongruent'].plot(kind='hist')

Out[21]: <matplotlib.axes._subplots.AxesSubplot at 0x7f60f63609e8>
```



The histograph of both studies shows clear distinction between congruent study and incongruent study. The congruent and incongruent histographs exhibit multiple peaks. However, on average the distribution of incongruent group is at higher time compare to congrent group.

(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.

Our sample size is less than 30 and we are unaware of population standard deviation. Therefore, as per this source (https://www.statisticshowto.datasciencecentral.com/probability-and-statistics/hypothesis-testing/t-score-vs-z-score/), it would be best to use T-test. Since, we are using same set of people to conduct two different experiments, it would be best to use "Paired T-test".

```
In [22]: # T-test
         from scipy.stats import ttest_rel

         ttest_rel(df['Incongruent'], df['Congruent'])

Out[22]: Ttest_relResult(statistic=8.020706944109957, pvalue=4.1030005857111781e-08)
```

p value is less than 0.05. Therefore, we can reject the null hypothesis. And we can safely conclude that average time it takes to read incongruent words is different than the time it takes to read congruent 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!

Our minds are conditioned to identify color than read letters first. Therefore, when it sees a name, different than associated color, it takes time to make this distinction.

Another example could incongruent numbers. For example, smaller number is written in bigger size while large number is written in the small size. This creates a delay in comprehension while reading these two numbers.