Test a Perceptual Phenomenon

January 23, 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?

- independent variable: word condition (congruent or incongruent).
- dependent variable: time to say the color of the ink in which the word is printed in seconds.

(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 hypothesis(H_0): the time required to say the written ink color of a congruent word equals the time required to say the written ink color of an incongruent word.
- alternative hypothesis(H_1): the time required to say the written ink color of a congruent word doesn't equal(less than) the time required to say the written ink color of an incongruent word.

$$H_0 : \mu_{congruent} = \mu_{incongruent}$$

(null hypothesis)

$$H_1 : \mu_{congruent} \neq \mu_{incongruent}$$

(alternative hypothesis) '

$\mu congruent$: population mean time for congruent word.

$\mu_{incongruent}$: population mean time for incongruent word.

(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 [21]: import pandas as pd
         import numpy as np
         import matplotlib.pyplot as plt
         import scipy.stats as stats
         % matplotlib inline

In [22]: df = pd.read_csv('stroopdata.csv')
         df.head()

Out[22]:    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 [23]: df.describe()

Out[23]:         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 [24]: #measures of central tendency:
         print('MEAN')
         print(df.mean())
         print()
         print('MEDIAN')
         print(df.median())

MEAN
Congruent      14.051125
Incongruent    22.015917
dtype: float64

MEDIAN
Congruent      14.3565
Incongruent    21.0175
dtype: float64


In [25]: #measures of variability:
         print('Standard Deviation')
         print(np.std(df))
```

```
print()
print('RANGE')
print(df.max()-df.min())
print()
print('Inter-quartile Range')
print('Congruent' + '      ' + str(df.Congruent.quantile(.75)-df.Congruent.quantile(.25
print('Incongruent' + '    ' + str(df.Incongruent.quantile(.75)-df.Incongruent.quantile
```

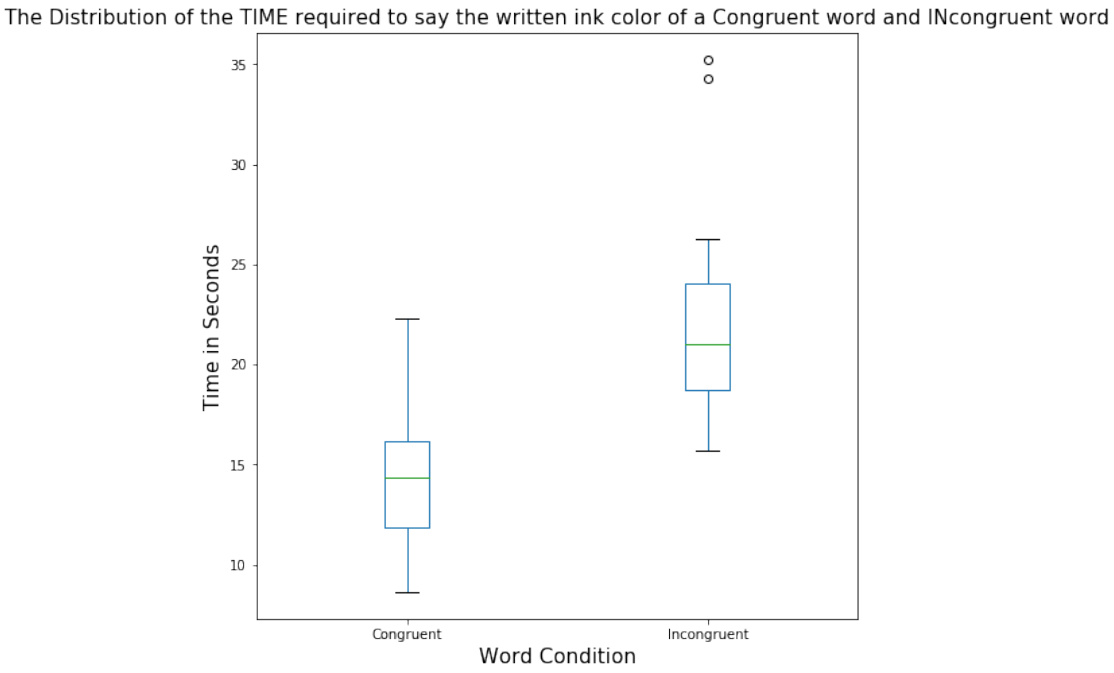
```
Standard Deviation
Congruent      3.484416
Incongruent    4.696055
dtype: float64

RANGE
Congruent      13.698
Incongruent    19.568
dtype: float64

Inter-quartile Range
Congruent      4.3054999999999986
Incongruent    5.33475
```

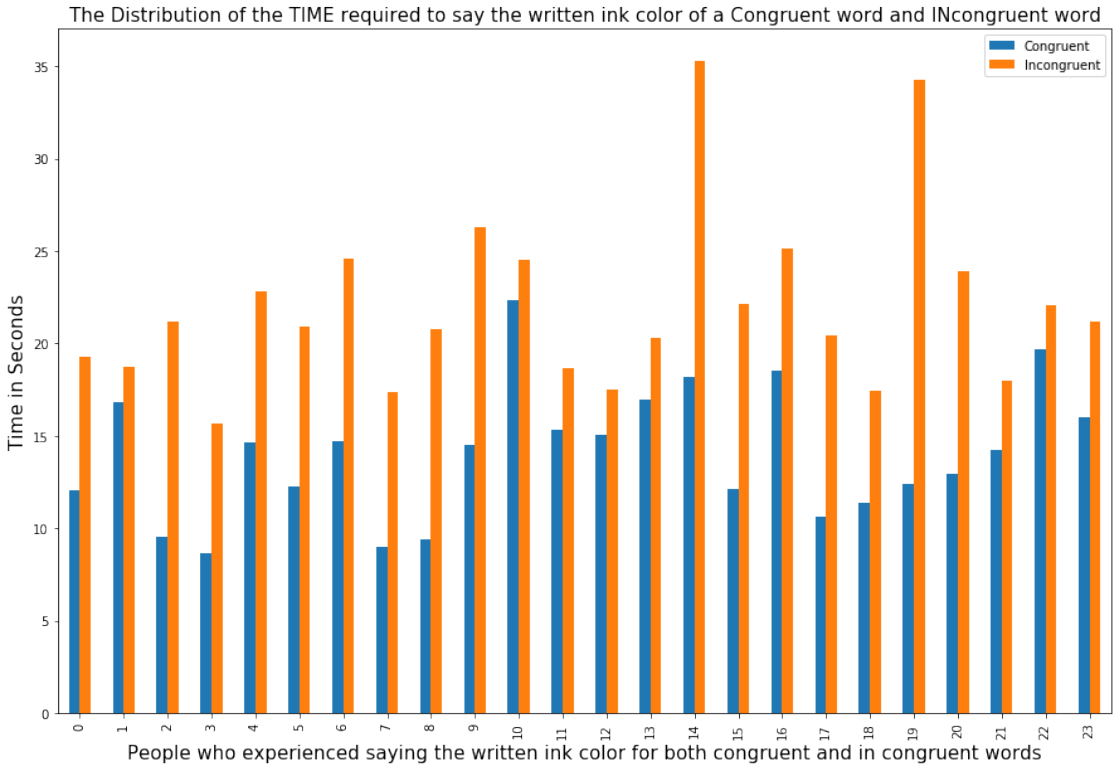
(4) Provide one or two visualizations that show the distribution of the sample data. Write one or two sentences noting what you observe about the plot or plots.

```
In [26]: df.plot(kind = 'box', figsize = (8,8));
         plt.title('The Distribution of the TIME required to say the written ink color of a Cong
         plt.xlabel('Word Condition', size = 15)
         plt.ylabel('Time in Seconds', size = 15);
```



- From the previous plot, we can educe that in general that time required to say the written ink color of a congruent word is less that the time required to say the written ink color of an incongruent word.
- in the box plot of the incongruent word, there are some outliers which means that some people take much time to say the written ink color of an incongruent word.
- The range of the incogruent box plot is wider that the range of the congruent box plot and this is because of the outliers.

```
In [27]: df.plot.bar(figsize = (15, 10));
         plt.title('The Distribution of the TIME required to say the written ink color of a Cong
         plt.xlabel('People who experienced saying the written ink color for both congruent and
         plt.ylabel('Time in Seconds', size = 15);
```



- From the previous plot, we can say that peolpe take more time to say the written ink color of the incongruent word than the congruent word.

(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.

- Now, I am going to perform some statiscal test and I am gonna choose the t-Test with dependent samples because this test here in our example here is performed two different times by the same people.

```
In [29]: stats.ttest_rel(df['Congruent'], df['Incongruent'])
```

```
Out[29]: Ttest_relResult(statistic=-8.020706944109957, pvalue=4.1030005857111781e-08)
```

- Supposing the confidence interval equals 95%, we will have alpha equals 0.05. From the previous result, we got p-value equals(4.1030005857111781e-08) and this value is less that the alpha, so we will reject the null hypothesis.
- This results totally matches my expectations because most of people will take more time to figure how to say the written ink color of a incongruent word.

(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!

```
In [ ]:
```