Test a Perceptual Phenomenon

October 15, 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?

```
In [ ]: Independent variable is Words condition whether congruent or incongruent.
        Dependent variable is time it takes to name the ink colour of a word.
```

--write answer here--

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

H0: Null hypothesis- Incongruent words has no effect on the time to read the ink colour of the word correctly. H1: Alternative hypothesis- Incongruent words has an effect or it increases the time to read the ink colour of the word correctly.

H0: ți=țc H1: ți>țc where ți is Population mean of incongruent values and țc is population means of congruent values.

Statisitcal test: I choose a confidence interval of 95% and paired one tail t-test. This test is chosen because each participant undergoes two tests and thus this test will tell if mean of incongruent values are statistically significantly different from those of congruent ones. Aplha level is 0.05.

Assumptions: a) I have chosen one-sided alternative t-test because my alternative hypothesis is that mean values of incongruent values are greater than mean values of congruent tests. b) t-test is chosen because of unknown standard deviation of the population and also sample size is less than 30 which approximates normal distribution thus limting the use of z-test.

(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 [2]: # Perform the analysis here
        import pandas as pd
        import numpy as np
        import scipy.stats as stats
        %matplotlib inline
        stroop=pd.read_csv('stroopdata.csv')
```

```python
print(stroop)
stroop.shape
x=stroop.describe()
print(x)
cong=stroop[["Congruent"]]
incong=stroop[["Incongruent"]]
print(np.mean(cong))
print(np.mean(incong))
print(np.std(cong))
print(np.std(incong))
```

```
     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
5       12.238        20.878
6       14.692        24.572
7        8.987        17.394
8        9.401        20.762
9       14.480        26.282
10      22.328        24.524
11      15.298        18.644
12      15.073        17.510
13      16.929        20.330
14      18.200        35.255
15      12.130        22.158
16      18.495        25.139
17      10.639        20.429
18      11.344        17.425
19      12.369        34.288
20      12.944        23.894
21      14.233        17.960
22      19.710        22.058
23      16.004        21.157
         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
Congruent     14.051125
dtype: float64
Incongruent     22.015917
```
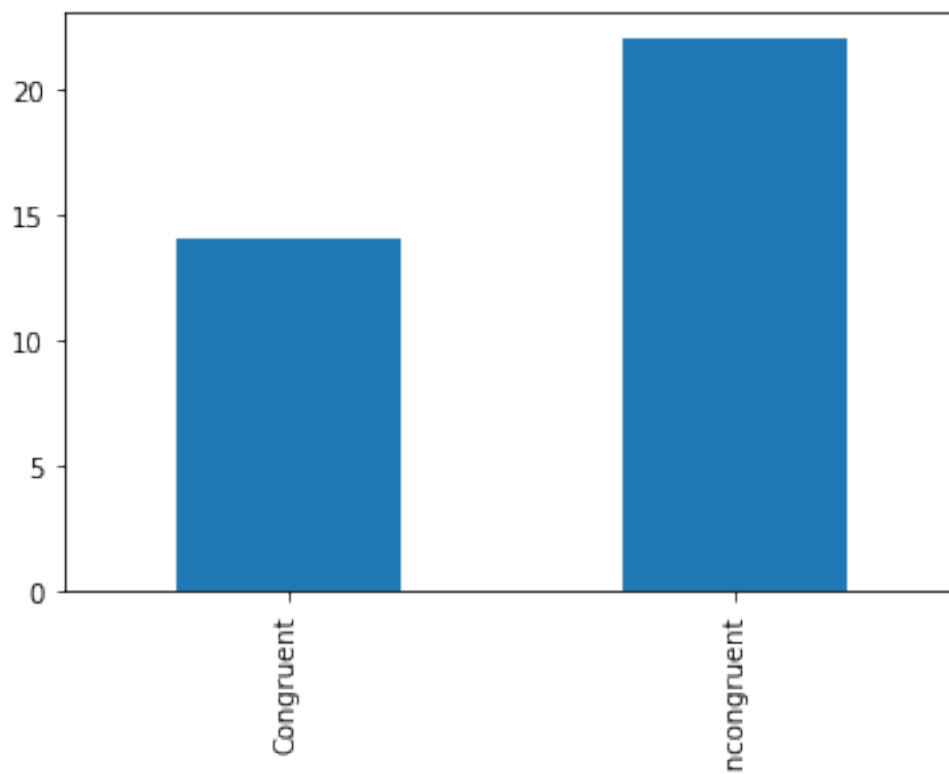
```
dtype: float64
Congruent    3.484416
dtype: float64
Incongruent    4.696055
dtype: float64
```
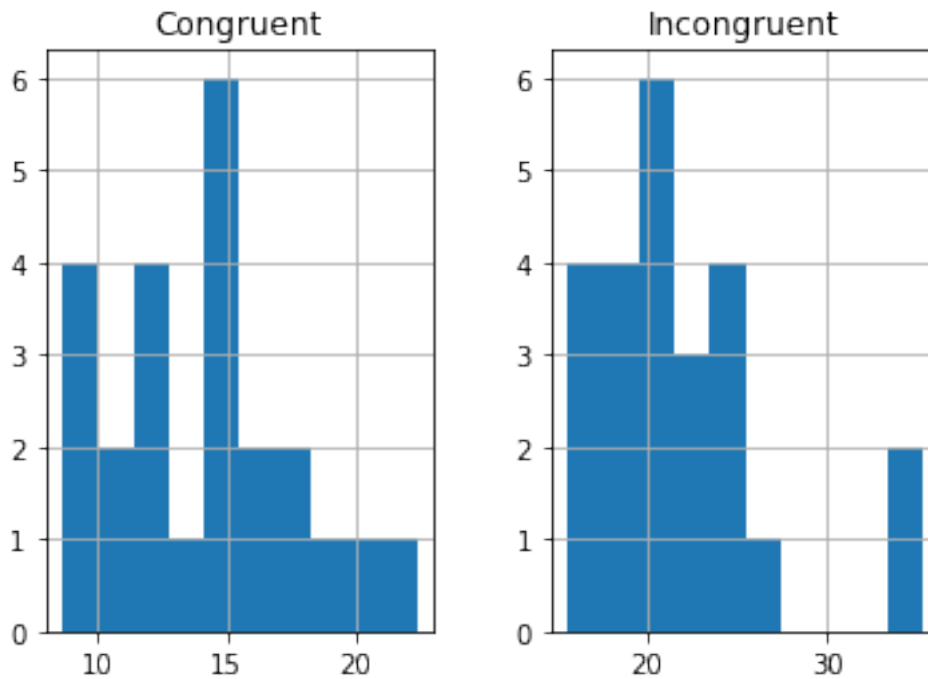
Mean and standard deviation of the Congruent test are 14.051125 and 3.4844 respectively. Mean and standard deviation of Incongruent test are 22.015917 and 4.6960 respectively. Thus the average time taken by the person for Congruent test is less than that of Incongruent test. Incongruent test requires a sharp perception of brain and it varies by individuals.

(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 [3]: # Build the visualizations here
        import matplotlib.pyplot as plt
        stroop=pd.read_csv('stroopdata.csv')
        stroop.shape
        cong=stroop[["Congruent"]]
        incong=stroop[["Incongruent"]]
        stroop.mean().plot('bar')
        stroop.hist()
```

```
Out[3]: array([[<matplotlib.axes._subplots.AxesSubplot object at 0x7ff75910d9b0>,
                <matplotlib.axes._subplots.AxesSubplot object at 0x7ff759086710>]], dtype=object
```

The bar chart verifies the descriptive stats performed in the previous question. Mean time taken by people for Incongruent test is high when compared to Congruent test. Histogram plots shows slightly postive skew in both cases and also an approximate normal distribution.

(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 [2]: import pandas as pd
        import numpy as np
        import scipy.stats as stats
        stroop=pd.read_csv('stroopdata.csv')
        print(stroop)
        cong=stroop[["Congruent"]]
        incong=stroop[["Incongruent"]]
        print(stats.ttest_rel(incong,cong))
```

	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

4

```
5       12.238       20.878
6       14.692       24.572
7        8.987       17.394
8        9.401       20.762
9       14.480       26.282
10      22.328       24.524
11      15.298       18.644
12      15.073       17.510
13      16.929       20.330
14      18.200       35.255
15      12.130       22.158
16      18.495       25.139
17      10.639       20.429
18      11.344       17.425
19      12.369       34.288
20      12.944       23.894
21      14.233       17.960
22      19.710       22.058
23      16.004       21.157
Ttest_relResult(statistic=array([ 8.02070694]), pvalue=array([  4.10300059e-08]))
```

T statistic = d-bar/(Sd/sqrt(n)) where d-bar is the average difference between Incongruent and congruent values, Sd is the standard deviation of the population and n is the sample size. Upon calculating, it is approximately 8.02. t-distribution with Degree of freedom 23 and at alpha 0.05, we get t as 1.714 and a very less p-value < 0.05 This gives suffcient clarity and confidence to reject the null hypothesis which states mean of Congruent and Incongruent time are equal. Thus the alternative hypothesis is considered concluding that the Incongruent mean values are greater than congruent values. The results match with the expectations.

```
In [ ]: Refrences
        1) https://classroom.udacity.com/courses/ud134-nd/lessons/4446458586/concepts/4614378711
        2)Design and Analysis of experiments, Douglas Montgomery textbook.
        3)Python lecture materials from data camp.
```