# Test a Perceptual Phenomenon

March 1, 2019

### 0.0.1 Analyzing the Stroop Effect

The Stroop effect is a phenomenon that occurs when you must say the color of a word but not the name of the word. For example, blue might be printed in red and you must say the color rather than the word.

While it might sound simple, the Stroop effect refers to the delayed reaction times when the color of the word doesn't match the name of the word. It's easier to say the color of a word if it matches the semantic meaning of the word. For example, if someone asked you to say the color of the word "black" that was also printed in black ink, it would be much easier to say the correct color than if it were printed in green ink.

The task demonstrates the effect that interference can have when it comes to reaction time. It was first described during the 1930s by American psychologist John Ridley Stroop for whom the phenomenon is named. His original paper describing the effect has become one of the most famous, as well as one of the most frequently cited, in the history of psychology. The effect has been replicated hundreds of times by other researchers.

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

**the independent variable is :** An independent variable is a variable that represents a quantity that is being manipulated in an experiment.a variable that changes or that we have control over. which is in the stroop effect (what the words say and the color of the words, congruent or incongruent).

**the dependant variable :** A dependent variable represents a quantity whose value depends on how the independent variable is manipulated.it's the variable being tested in an experiment. which is in the stroop effect (the reading time).

(2) What is an appropriate set of hypotheses for this task? Specify your null and alternative hypotheses, and clearly define any notation used.

$$H_0 : \mu_1 - \mu_2 = 0$$

$$H_1 : \mu_1 - \mu_2 \neq 0$$

$H_0$ is the null hypotheses : there is no difference between the population means of congruent and incongruent times. $H_1$ is the alternative hypotheses : there is a difference between the population means of congruent and incongruent times.

$\mu_1$ is the average of congruent times. $\mu_2$ is the average of incongruent times.

i have chosen the previous null and alternative hypotheses because we are trying to find out if the difference between the average congruent and incongruent reading time is statistically significant.

the test that will be used is a paired t-test which is a dependent test, because the same group of units has been tested twice.

(3) Report some descriptive statistics regarding this dataset. Include at least one measure of central tendency and at least one measure of variability.

```
In [1]: # Perform the analysis here

        import pandas as pd
        import matplotlib.pyplot as plt

        %matplotlib inline

        df = pd.read_csv('stroopdata.csv')
        df.head()
```

Out[1]:

| | 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 [2]: df.shape
```

Out[2]: (24, 2)

```
In [3]: df
```

Out[3]:

| | 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
```

```
In [4]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 24 entries, 0 to 23
Data columns (total 2 columns):
Congruent      24 non-null float64
Incongruent    24 non-null float64
dtypes: float64(2)
memory usage: 464.0 bytes
```

Measure of central tendency :

```
In [5]: print('the average Congruent reading time is : {}'.format(df['Congruent'].mean()))
        print('the average Incongruent reading time is : {}'.format(df['Incongruent'].mean()))
```

```
the average Congruent reading time is : 14.051124999999999
the average Incongruent reading time is : 22.015916666666666
```

Measure of variability :

```
In [6]: print('the standard deviation of Congruent reading time is : {}'.format(df['Congruent'].
        print('the standard deviation of Incongruent reading time is : {}'.format(df['Incongruen
```

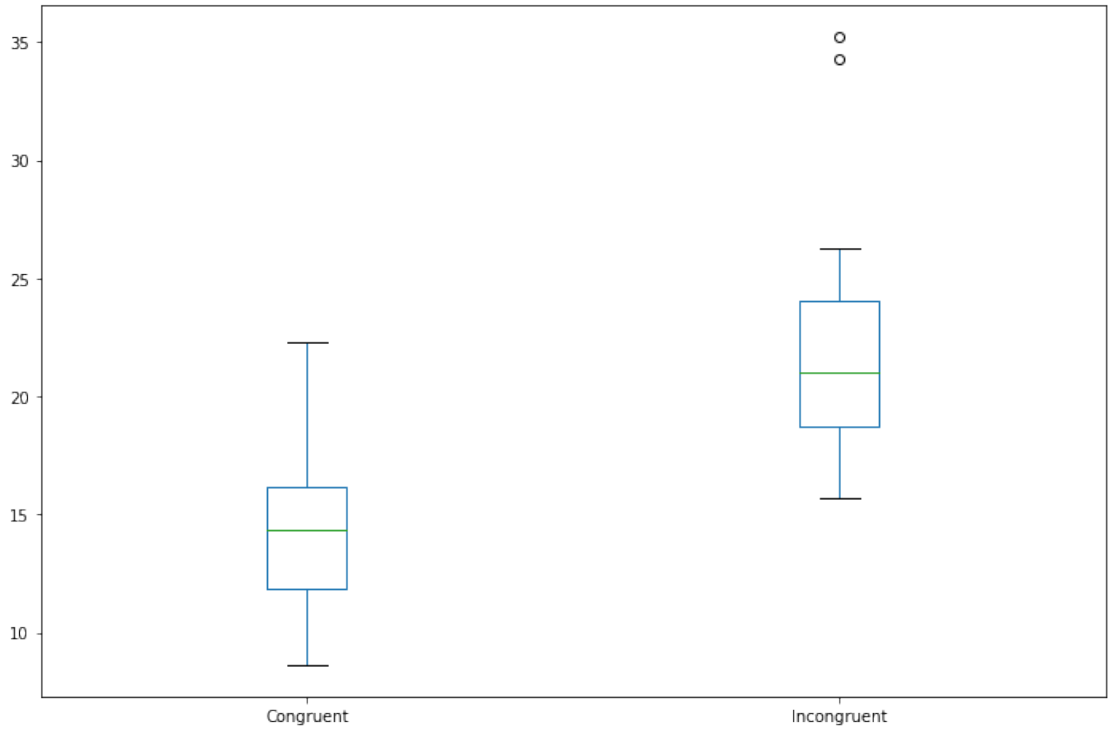
```
the standard deviation of Congruent reading time is : 3.5593579576451955
the standard deviation of Incongruent reading time is : 4.797057122469138
```

```
In [7]: 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
```

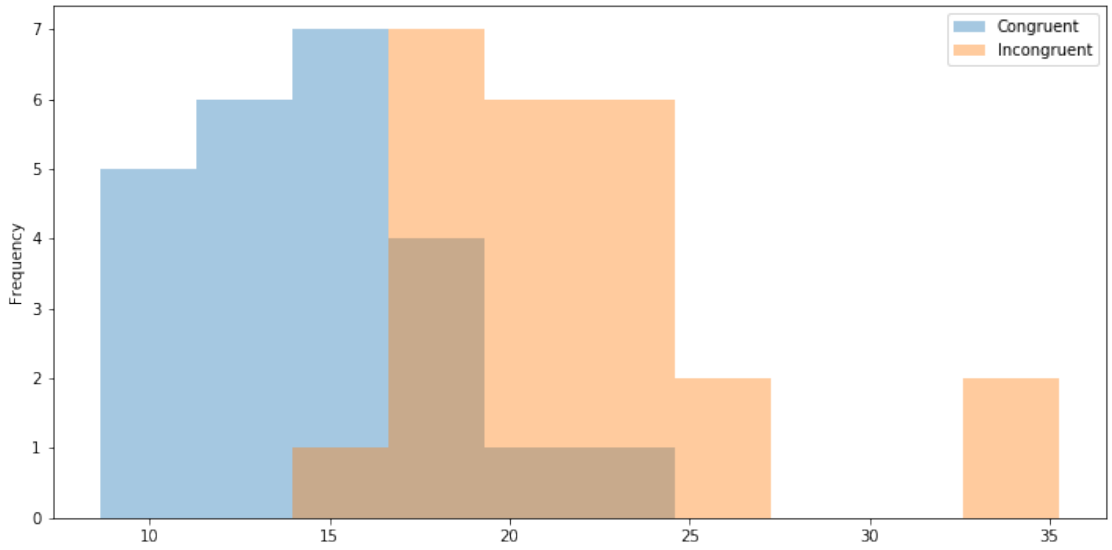
(4) Provide one or two visualizations that show the distribution of the sample data. Write one or two sentences noting what you observe about the plot or plots.

```
In [8]: # Build the visualizations here
        df.plot(kind='box',figsize=(12,8));
```




After observing and comparing the box plots of both congruent and incongruent reading times,we can see that on average the incongruent reading times are higher, and we can see that the incongruent reading times also have two outliers at about 35 while the congruent reading times don't have any.

```
In [9]: df.plot(kind='hist',alpha=0.4,figsize=(12,6));
```




from the histogram above,we can derive almost the same result which are on average the incongruent reading times are higher, and incongruent reading times have outliers at about 35 while the congruent reading times don't have any.

(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?

The confidence interval : 95% the threshold (alpha) : 0.05 or 5%

```
In [10]: # Perform the statistical test here

         from scipy import stats

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

```
Out[10]: Ttest_relResult(statistic=-8.020706944109957, pvalue=4.1030005857111781e-08)
```

the p-value $\approx$ 0.00000004 the p-value is much smaller than our $\alpha$ (0.05), so we reject the null hypotheses. this confirms that our previous result is statistically significant.

### 0.1 Refrences :

https://www.khanacademy.org/math/pre-algebra/pre-algebra-equations-expressions/pre-algebra-dependent-independent/a/dependent-and-independent-variables-review https://en.wikipedia.org/wiki/Student%27s_t-test#Dependent_t-test https://docs.scipy.org/doc/scipy-0.14.0/reference/generated/scipy.stats.ttest_rel.html