Test a Perceptual Phenomenon

March 23, 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 and submit in the final section of this lesson.

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

In the Stroob Effect, the independent variable is whether the pairs are congruent or incongruent. The dependent variable is the time taken to name the colors.

(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.

In performing this experiemnt, we test whether there's a significant difference in completion time between congruent and incongruent sets of colors. A sample of the population was taken and administered the Stroob test. We need to determine whether theres a significant difference between the Congruent sample mean (C) and the Incongruent sample mean (I) Our Null Hypothesis(H0) is there isn't any significant difference between the congruent(C) completion time and the incongruent(I) completion time. Our Alternative Hypothesis (Ha) is there is a significant difference between the congruent(C) completion time and incongruent(I) completion time.

H0: uC = uI

HA: uC != uI

Since the same sample was repeatedly measured, with and without the independent variable, a paired T-test will be used to test whether there is a statistically significant difference between the mean completion times of the congruent and incongruent data sets. In using this test, the following assumptions are met: the dependent variable is measured by interval, a random sampling was taken from a population, the before and after measurements are paired, and the scores are normally distributed.

(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]: # Import packages
        import pandas as pd
        import scipy.stats as stats
        import numpy as np
```

```
        import matplotlib as mpl
        import matplotlib.pyplot as plt
        %matplotlib inline
```

```
In [3]: # load dataframe
        data = pd.read_csv('stroopdata.csv')
        #check mean
        data.mean()
```

```
Out[3]: Congruent      14.051125
        Incongruent    22.015917
        dtype: float64
```

```
In [4]: data['Congruent'].values.std(ddof=1)
```

```
Out[4]: 3.5593579576451955
```

```
In [5]: data['Incongruent'].values.std(ddof=1)
```

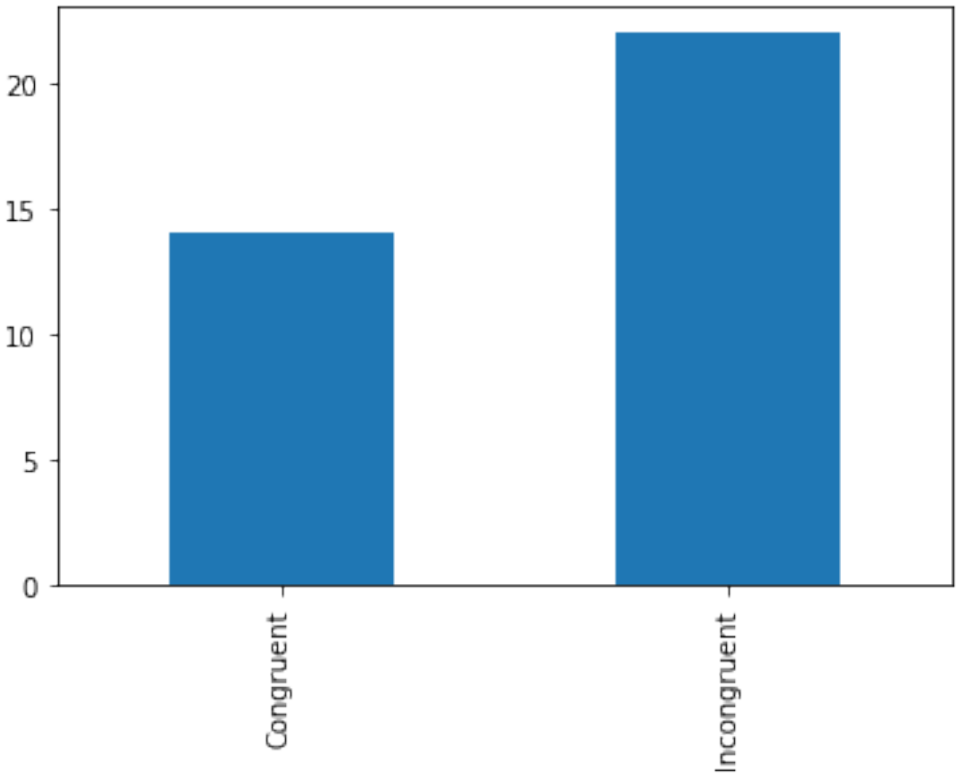
```
Out[5]: 4.7970571224691376
```

There seems to be a notable difference between the mean of the congruent and incongruent times. Additionally, the incongruent times have a greater variance, while the congruent times trend towards the mean.

(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.

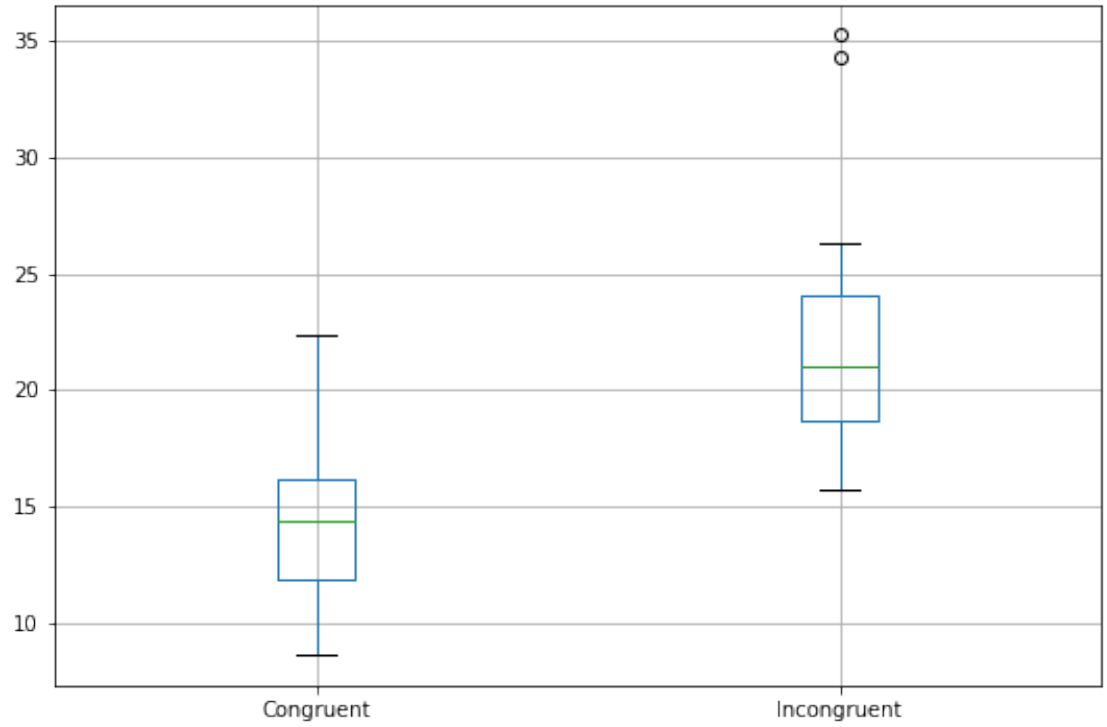
```
In [6]: # Set columns to lists to use in ttest function
        cong = data['Congruent'].values.tolist()
        inco = data['Incongruent'].values.tolist()

        # Plot the means
        data.mean().plot('bar')
```

```
Out[6]: <matplotlib.axes._subplots.AxesSubplot at 0x7f7df20b2dd8>
```




```
In [7]: data.boxplot()
```

```
Out[7]: <matplotlib.axes._subplots.AxesSubplot at 0x7f7df2067a20>
```



There is a noticible difference between the Congruent time lapse and the Incongruent time lapse. A Paired T-test will determine whether this difference is statistically significant or not.

(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 [10]: # Perform ttest
         stats.ttest_rel(cong,inco)
```

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

The T-test returns a pvalue < .01, which means we reject H0. This means there is a significant difference in time between identifying congruent and incongruent tests. While 95% is a common confidence level, the pvalue is so low we can asser this with 99% confidence level.

(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!

I suspect this difference is caused by the speed the brain identifies words as opposed to colors. I think a similiar test would be if you used directions, say UP DOWN LEFT RIGHT, and had them flash onto the screen congruently and incongruently.

Sources: https://stackoverflow.com/questions/25140998/pandas-compute-mean-or-std-standard-deviation-over-entire-dataframe

http://www.scipy-lectures.org/packages/statistics/index.html

https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.ttest_rel.html

https://en.wikipedia.org/wiki/Stroop_effect#Theories