Project01

July 16, 2015

```
In [1]: #My Results
        t1 = 18.99
        t2 = 21.152
```

1 What is our independent variable? What is our dependent variable?

The independent variable is which set of words (congruent or incongruent) is shown. The dependent variable is the time it takes to read the words.

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

A two-tailed test could be used with a null hypothesis that the reading time is the same for both sets, and an alternative hypothesis that the reading time is significantly different. A paired t-test would be useful since the timings are paired for the same person.

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

```
In [2]: import pandas as pd
        df = pd.read_csv('stroopdata.csv')
        print(df.mean())
        print(df.std())
```

```
Congruent      14.051125
Incongruent    22.015917
dtype: float64
Congruent      3.559358
Incongruent    4.797057
dtype: float64
```

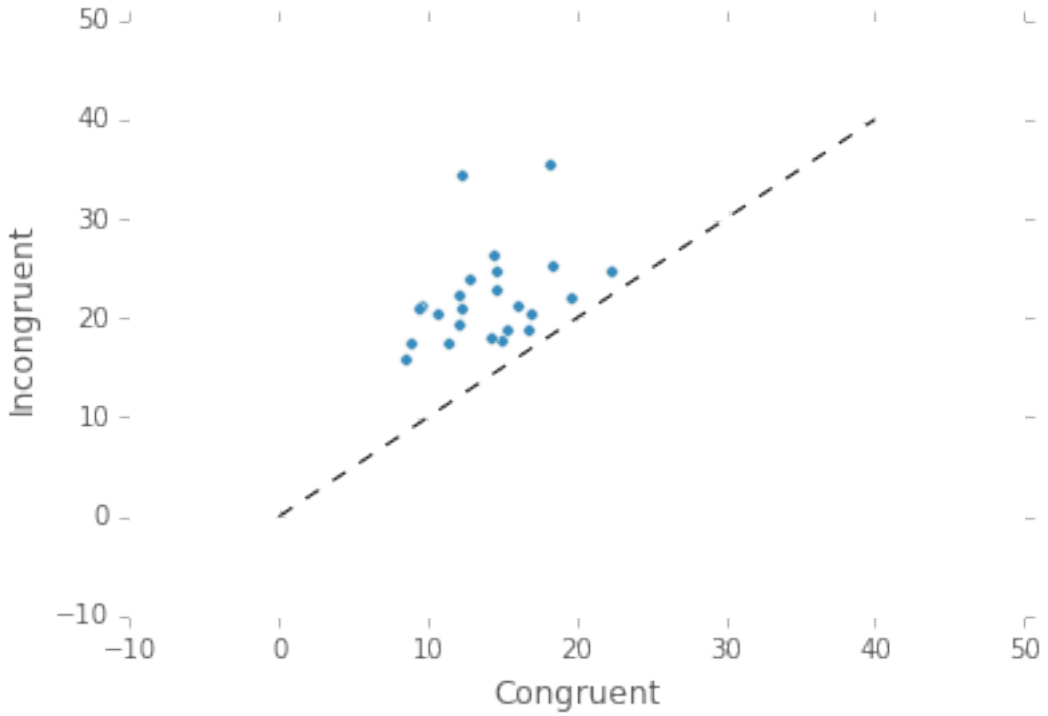
As seen above, the average time is 14.05 seconds and 22.02 seconds for congruent and incongruent (respectively). The standard deviations are 3.56 and 4.80.

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]: import matplotlib.pyplot as plt
        plt.style.use('ggplot')
        %matplotlib inline

        fig,ax = plt.subplots(1)
        df.plot(x='Congruent', y='Incongruent', kind='scatter', ax=ax)
        ax.plot([0,40],[0,40], ls='--', color='black')
```

```
Out[3]: [<matplotlib.lines.Line2D at 0x7f0a1a6a99e8>]
```



It looks like reading the incongruent set of words always took more time than reading the congruent set of words. I've drawn a diagonal line on the scatterplot to show where equal values would occur.

5 Now, perform the statistical test and report your results. What is your 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 [4]: import numpy as np
        diff = df['Incongruent']-df['Congruent']
```

```
diff_stderror = diff.std()/np.sqrt(len(diff))
t_stat = (diff.mean()-0)/diff_stderror
print('The T-Statistic is {}'.format(t_stat))
```

```
The T-Statistic is 8.020706944109953
```

There are 23 degrees of freedom (24 observations). Using a confidence level of 95%, the critical test statistic is 2.069. The t-statistic greatly exceeds this, so I reject the null hypothesis and conclude that there is a significant difference in reading speed of the two datasets (decreased speed for incongruent words). This matches my expectations after trying the test myself and viewing the scatterplot.

6 Other sources:

- http://pandas.pydata.org/pandas-docs/stable/generated/pandas.DataFrame.std.html - To confirm this returns the unbiased standard deviation
- https://en.wikipedia.org/wiki/Stroop_effect - Background Information

In [ ]: