StroopTest

November 8, 2015

1 Statistics: The Science of Decisions

```
In [1]: from pandas import *
        dataFrame = read_csv('stroopdata.csv')
```

```
In [2]: dataFrame
```

```
Out[2]:     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
```

Variables The independent variable in this case is the

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

Null and alternative hypotheses The null hypothesis is that population mean of congruent times is equal to the population mean of incongruent times. - H_0: $\mu_C = \mu_I$ (where μ_C is the congruent population mean and μ_I is the incongruent population mean.)

The alternative hypothesis is that population mean of congruent times is *not* equal to the population mean of incongruent times. - H_1: $\mu_C \neq \mu_I$

Statistical test rationale Using sample data from congruent and incongruent times, we will conduct a *two- tailed, two-sample dependent t-test* of the null hypothesis. We will making the following assumptions about the data

- Participants have recorded a congruent and subsequent incongruent time.
- Population of congruent and incongruent times are iid normal.
- Variances of groups *are* assumed to be equal.
- Data distributions that are roughly symmetric/mound (Gaussian) shaped and not skewed.

For this analysis, we will use a signficance value of $\alpha = 0.05$.

Measures of central tendency and variance

```
In [3]: #Calculate the mean time to complete the congruent and incongruent exercise
        print "The mean time taken for the congruent condition is =", dataFrame["Congruent"].mean()
        print "The mean time taken for the incongruent condition is =", dataFrame["Incongruent"].mean()

        #Calculate the  the congruent and incongruent variances
        print "The congruent time variance =", dataFrame["Congruent"].var()
        print "The incongruent time variance =", dataFrame["Incongruent"].var()
```

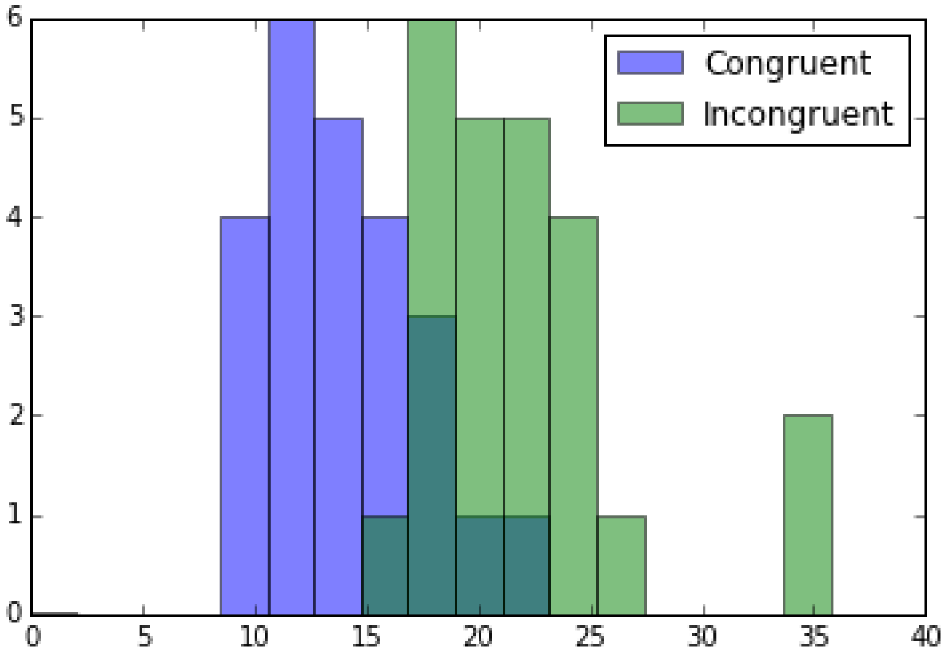
```
The mean time taken for the congruent condition is = 14.051125
The mean time taken for the incongruent condition is = 22.0159166667
The congruent time variance = 12.6690290707
The incongruent time variance = 23.0117570362
```

Exploring the data with visualisations

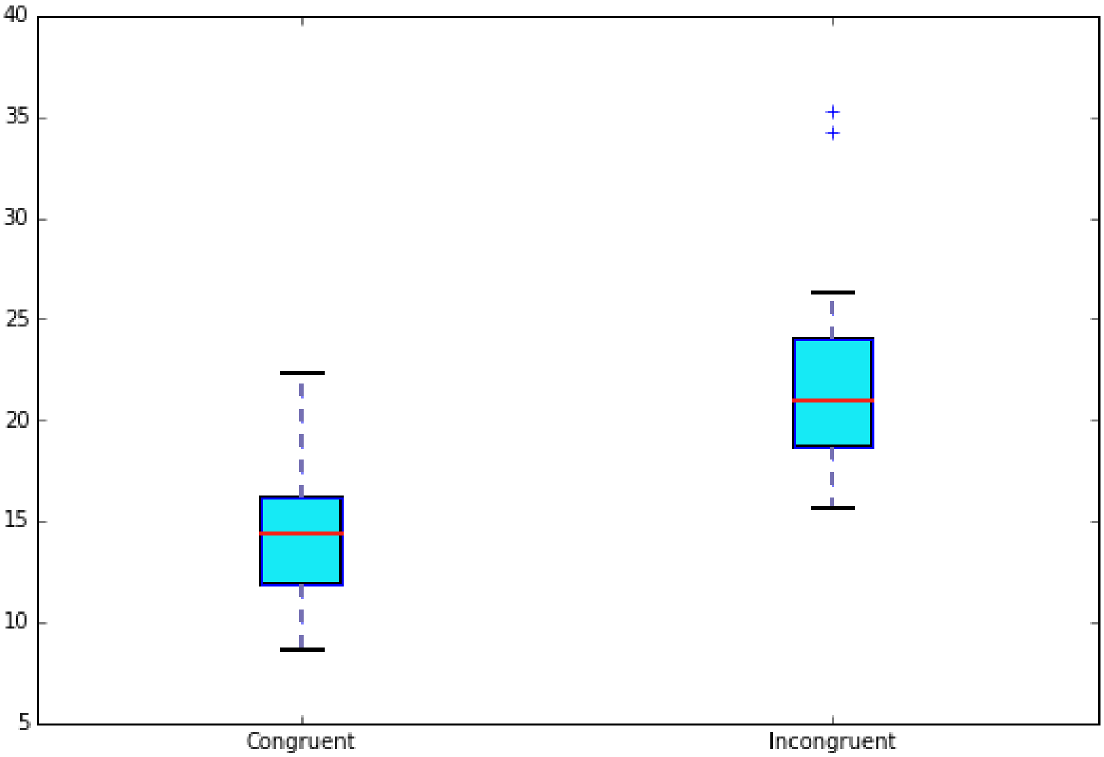
```
In [9]: #Histogram of congruent and incongruent
        #source: http://stackoverflow.com/questions/6871201/plot-two-histograms-at-the-same-time-with-m

        import matplotlib as mpl
        mpl.use('agg')
        import matplotlib.pyplot as plt
        %matplotlib inline

        import numpy
        from pandas import *


        dataFrame = read_csv('stroopdata.csv')

        x = dataFrame["Congruent"]
        y = dataFrame["Incongruent"]

        bins = numpy.linspace(0, 40, 20)

        plt.hist(x, bins, alpha=0.5, label='Congruent')
        plt.hist(y, bins, alpha=0.5, label='Incongruent')
        plt.legend(loc='upper right')
        plt.show()

        #Box plot of the congruent and incongruent times
        #Source: http://blog.bharatbhole.com/creating-boxplots-with-matplotlib/
```

```
dataFrame = read_csv('stroopdata.csv')

plotData = [dataFrame["Congruent"], dataFrame["Incongruent"]]

# Create a figure instance
fig = plt.figure(1, figsize=(9, 6))

# Create an axes instance
ax = fig.add_subplot(111)

# Create the boxplot
bp = ax.boxplot(plotData)

# Save the figure
fig.savefig('fig1.png', bbox_inches='tight')

## add patch_artist=True option to ax.boxplot()
## to get fill color
bp = ax.boxplot(plotData, patch_artist=True)

## change outline color, fill color and linewidth of the boxes
for box in bp['boxes']:
    # change outline color
    box.set(facecolor = '#16EAF5', linewidth=2)

## change color and linewidth of the whiskers
for whisker in bp['whiskers']:
    whisker.set(color='#7570b3', linewidth=2)

## change color and linewidth of the caps
for cap in bp['caps']:
    cap.set(linewidth=2)

## change color and linewidth of the medians
for median in bp['medians']:
    median.set(color='#F62217', linewidth=2)

ax.set_xticklabels(['Congruent', 'Incongruent'])
```

```
C:\Anaconda\lib\site-packages\matplotlib\__init__.py:1318: UserWarning:  This call to matplotlib.use() ha
because the backend has already been chosen;
matplotlib.use() must be called *before* pylab, matplotlib.pyplot,
or matplotlib.backends is imported for the first time.

  warnings.warn(_use_error_msg)
```


6
5
4
3
2
1
0
0
5
10
15
20
25
30
35
40
Congruent
Incongruent

```
Out[9]: [<matplotlib.text.Text at 0x17da3080>, <matplotlib.text.Text at 0x17db87b8>]
```


40
35
30
25
20
15
10
5
Congruent
Incongruent

Performing paired t-test Since $\alpha = 0.05$, we have a confidence level of 95% for our hypothesis test The critical t-values of the test with $df = 22$ are 2.074 and -2.074.

If the t-statistic is in the critical region, i.e. the t-statistic > 2.074 ot t-statistic < -2.074, we can reject the null hypothesis $\mu_C = \mu_I$, and accept the alternative hypothesis $\mu_C \neq \mu_I$. If the t-statistic is between -2.074 and 2.074, we fail to reject the null hypothesis.

```
In [10]: import scipy
         from scipy import stats

         #Calculate t-statistic
         print 'The t-statistic for the test = %6.3f pvalue = %s' %  scipy.stats.ttest_rel(x, y)
         print 'with df = 22'
```

```
The t-statistic for the test = -8.021 pvalue = 4.10300058571e-08
with df = 22
```

Conclusions The calculated t-statistic for the congruent and incongruent samples is -8.021. This is less than the critical t-value -8.021, therefore we can reject the null hypothesis with more than 95% confidence.

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?
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 [ ]:
```

```
In [ ]:
```

```
In [ ]:
```