

P1-v2

November 5, 2016

1 Stroop Effect

1.1 Background Information

In a Stroop task, participants are presented with a list of words, with each word displayed in a color of ink. The participant's task is to say out loud the color of the ink in which the word is printed. The task has two conditions: a congruent words condition, and an incongruent words condition. In the congruent words condition, the words being displayed are color words whose names match the colors in which they are printed: for example RED, BLUE. In the incongruent words condition, the words displayed are color words whose names do not match the colors in which they are printed: for example PURPLE, ORANGE. In each case, we measure the time it takes to name the ink colors in equally-sized lists. Each participant will go through and record a time from each condition.

1.2 Questions For Investigation

As a general note, be sure to keep a record of any resources that you use or refer to in the creation of your project. You will need to report your sources as part of the project submission.

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

The independent variable is the task conditions: congruent words condition or incongruent words condition.

The dependent variable is the recorded time from each condition.

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

The Null hypothesis would be: there is no significant difference of time from each condition. The Alternative Hypothesis would be: there is a significant difference of time from each condition.

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

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

With μ_1 the mean time for the congruent condition and μ_2 the mean time for the incongruent condition I would propose a 2 tail t-test. 2 tail because in the hypothesis I proposed, there is no indication whether the time for one condition is increased in comparison to the other. It will be a t-test because, the population parameters are unknown.

It is paired samples (same participants but 2 conditions) and therefore I focus on the time difference between the two test of each participant.

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

The mean time for the sample in congruent condition is 14.05s. The mean time for the sample in incongruent condition is 22.02s. The median time value for the congruent condition is 14.36s whereas the median time value for the incongruent condition is 21.02s.

With the bin size of 2s, the mode for the congruent condition is 14-16s whereas it 20-22s for the incongruent condition.

The sample standard deviation of the time in congruent condition is 3.48s whereas in incongruent condition it is equal to 4.70s. The population standard deviation estimation of the time for the congruent condition is 3.56s and it is 4.80s in incongruent condition.

The range for the congruent condition is 13.70s whereas it is 19.57s for the incongruent condition. The interquartile range is 4.31s for the congruent test whereas it is 5.33s for the incongruent condition.

```
In [1]: # import all necessary packages and functions.
import csv
from datetime import datetime
import numpy as np
import pandas as pd
import math
%matplotlib inline

# file locations
file_in = 'stroopdata.csv'

sample_data = pd.read_csv(file_in)

#display(sample_data.head())

a=sample_data["Congruent"]
b=sample_data["Incongruent"]

print "Mean, Population Standard Deviationestimation, Interquartile range,"
print a.mean(), a.std(),a.quantile(0.75)-a.quantile(0.25), a.max()-a.min(),
print b.mean(), b.std(),b.quantile(0.75)-b.quantile(0.25), b.max()-b.min(),

print "Sample Standard Deviation"
tmp1=0.
meana=a.mean()
tmp2=0.
meanb=b.mean()
for i in range(0,len(a)):
    tmp1+=(a[i]-meana)**2
    tmp2+=(b[i]-meanb)**2
ssd1=math.sqrt(tmp1/len(a))
```

```
sstd2=math.sqrt(tmp2/len(b))
```

```
print sstd1,sstd2
```

```
Mean, Population Standard Deviation, Interquartile range, Range, Median
14.051125 3.55935795765 4.3055 13.698 14.3565
22.0159166667 4.79705712247 5.33475 19.568 21.0175
Sample Standard Deviation
3.48441571277 4.69605513451
```

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.

On the histograms, it can be seen that the distribution for the congruent and incongruent condition looks relatively normal. On the box plot, it can be seen that there is no outliers for the congruent condition whereas there are 2 for the incongruent condition.

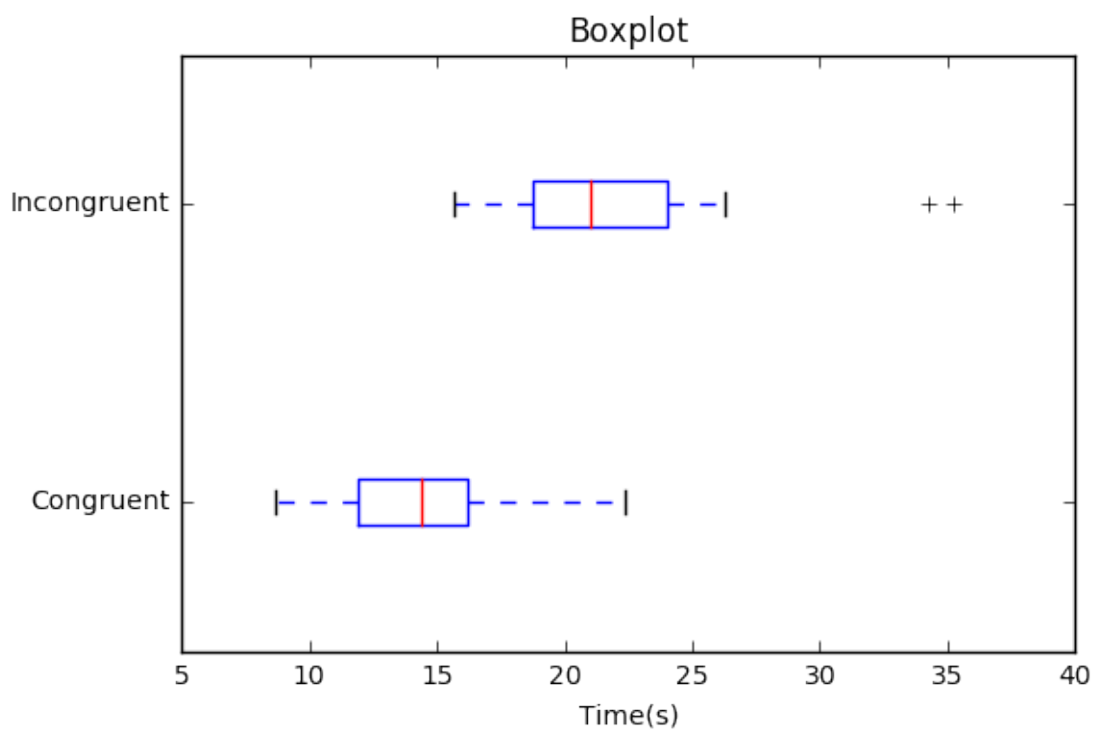
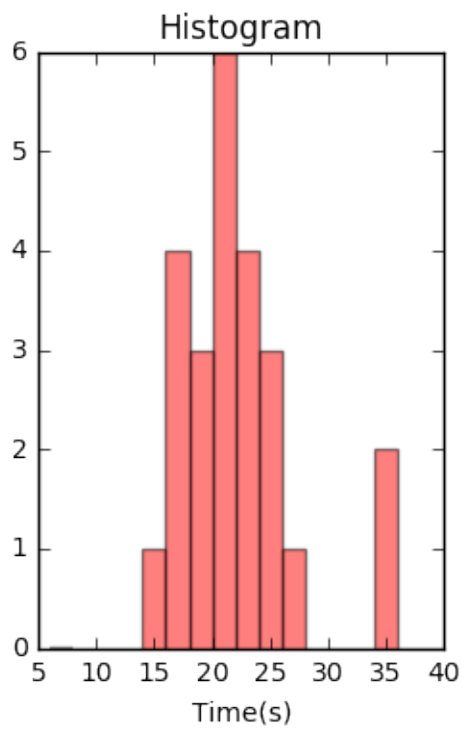
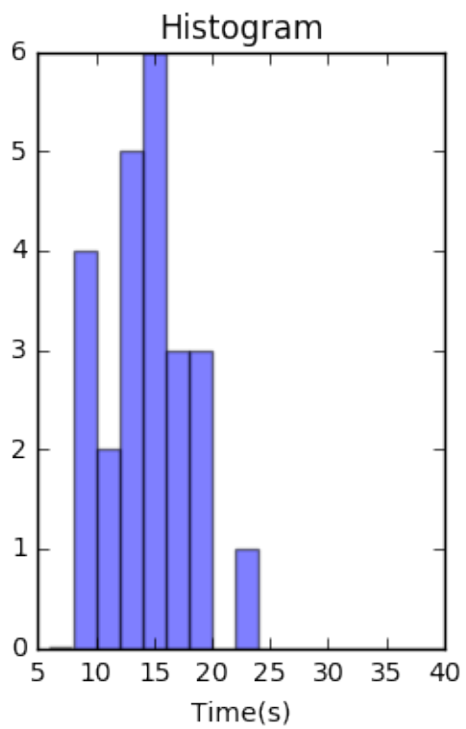
```
In [2]: import matplotlib.pyplot as plt
```

```
bins = np.linspace(6, 40, 18)
```

```
fig1, axarr1 =plt.subplots(1,2)
n, bins, patches = axarr1[0].hist([a,],bins, alpha=0.5)
n, bins, patches = axarr1[1].hist([b,],bins, alpha=0.5,facecolor='red')
axarr1[0].set_xlabel("Time(s)")
axarr1[1].set_xlabel("Time(s)")
axarr1[0].set_title("Histogram")
axarr1[1].set_title("Histogram")
```

```
fig2, axarr2 =plt.subplots(1,1)
tmp=axarr2.boxplot([a,b], vert=False)
axarr2.set_yticklabels(["Congruent", "Incongruent"])
axarr2.set_xlabel("Time(s)")
axarr2.set_title("Boxplot")
```

```
Out[2]: <matplotlib.text.Text at 0x9ea9630>
```



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?

The size of the sample is $n = 24$. The number of degree of freedom is then $df = n - 1$ which gives $df = 23$. The α -level I chose for the confidence level is $\alpha = 0.05$. From the t-table, $t_{critical} = 2.069$. The value for the t-test is $t_{value} = -8.02$. With such value, I reject the null the null hypothesis. The condition of the test (congruent or incongruent) has a significant impact on the result to the test. This result match my expectation. However, it would be possible to change the test to a one-tail t-test with the following null hypothesis: the time for the incongruent condition is below or equal to the time for the congruent condition.

```
In [3]: # size of the dataset
n=len(a)
print "n=",n

#Difference of dataset
d=a-b

# Mean of Sample of Difference
m=d.mean()
print "mean=",m

# Population standard deviation estimation
s=d.std()
print "s=",s

# tvalue calculation
tvalue=m/(s/math.sqrt(n))
print "tvalue=",tvalue

n= 24
mean= -7.96479166667
s= 4.86482691036
tvalue= -8.02070694411
```

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 think the effect observed is due to the activation of different zone of the brain: the one responsible for the reading capability and the one managing the vision color. When the information coming from are incoherent, the brain need to sort out which information is the right one with respect to the exercise.

From wikipedia, I read that the same kind of test could be done with Emotions, Directions and Numbers