P1. Stroop Test

Tiantian Lin

Udacity NanoDegree in Data Analyst

October 19, 2016

In a Stroop task, participants are presented with a list of words, with each word displayed in a color of ink. The participants 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.

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

Answer 1. Here, the independent variable can be viewed as a dummy variable 0 or 1, representing whether the participants' task is congruent or incongruent. The dependent variable is the time it takes to complete the task.

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

Answer 2. Here the set of hypothesis is

$$H_0 : \mu_{congruent} - \mu_{incongruent} = 0$$
$$H_A : \mu_{congruent} - \mu_{incongruent} \neq 0$$

The null hypothesis states that the mean of congruent and incongruent group are indifferent, while the alternative hypothesis states the opposite, i.e. the congruent and incongruent group are of different mean.

Two-tailed dependent t-test for paired sample should be suitable to test this effect. It is majorly because the experiment is carried out repeatedly on the same participant. Meanwhile, the scale of measurement is interval, the data is collected from random sampling. However, due to the relative small number of sampling, we need to assume that data points are being drawn from populations that are normally distributed, so that we can use the t-test. On the other hand, the alternative hypothesis we are interested in to reject has critical region falling in each tail, so I choose two-tailed test.

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

Answer 3. The sample mean and sample variance of the congruent and incongruent group are (14.05, 12.67) and (22.02, 23.01) respectively.

Exercise 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.
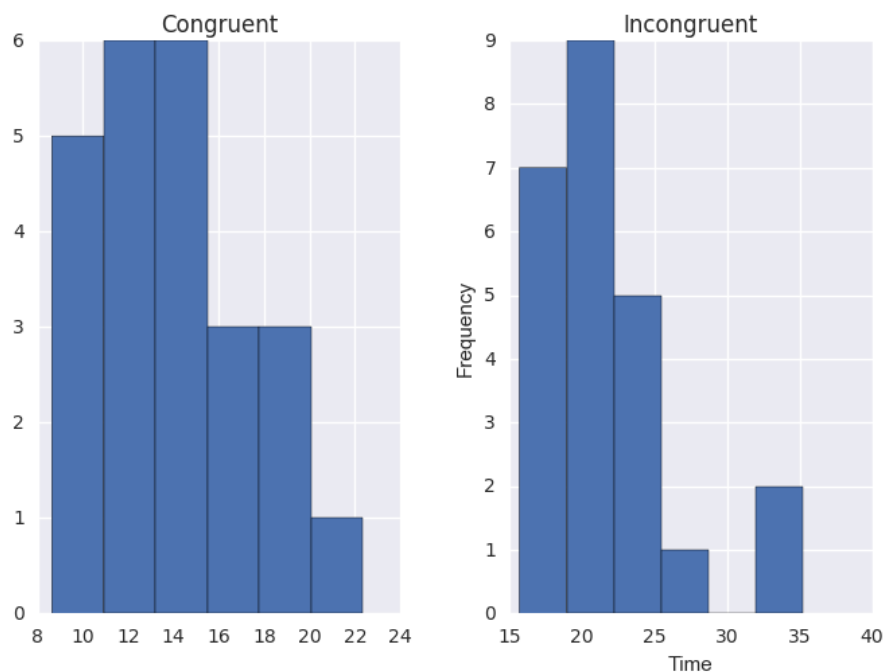
Answer 4. With the following code, I read the csv file into Python notebook to plot the histogram and boxplot.
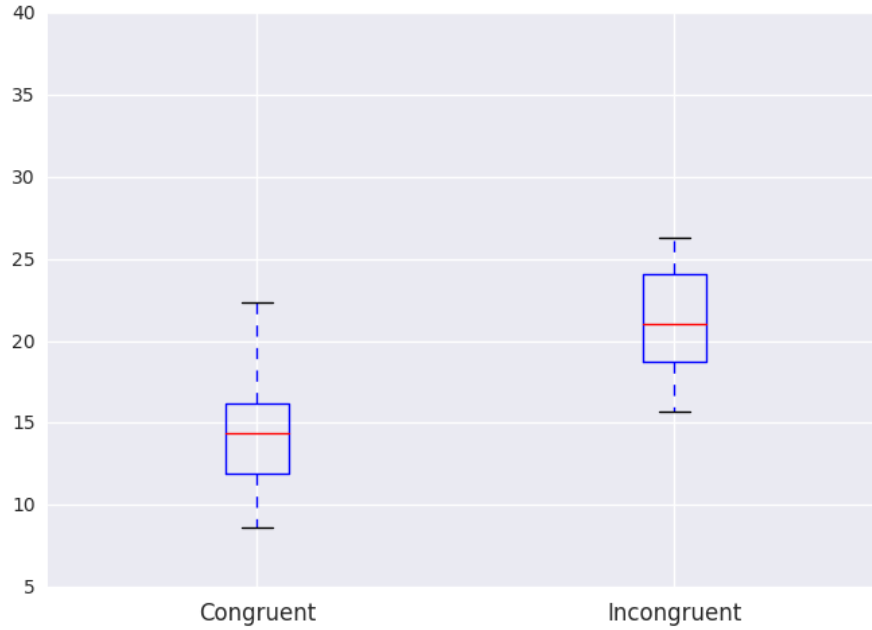
```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn

df=pd.read_csv("stroopdata.csv")

df.hist(bins='auto')
plt.xlabel("Time")
plt.ylabel("Frequency")
plt.legend(loc='upper right')
plt.show()

df.boxplot()
plt.show()
```

It is well noted from above visualization that the mode and mean of the incongruent group is greater than the congruent group. It shows, on average, it takes more time for participants to finish the incongruent task.

Exercise 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?

Answer 5. I set the hypothesis as

$$H_0 : \mu_{congruent} - \mu_{incongruent} = 0$$
$$H_A : \mu_{congruent} - \mu_{incongruent} \neq 0$$

Then I choose the confidence level at 95%. The degree of freedom is 23. It is a two tail test, so the critical t-stat is ± 2.069.

With the calculation (as follow), I calculated the t value at -8.02, which falls in the critical region. As a result, I reject the null hypothesis at a confidence level of 95%.

	A	B	C	D	E	F
1	Congruent	Incongruent	difference			
2	12.079	19.278	-7.199	-7.964792	<-- mean difference	
3	16.791	18.741	-1.95	4.8648269	<-- standard deviation S	
4	9.564	21.214	-11.65	0.9930286	<-- SEM	
5	8.63	15.687	-7.057			
6	14.669	22.803	-8.134	-8.020707	<-- t value	
7	12.238	20.878	-8.64			
8	14.692	24.572	-9.88			
9	8.987	17.394	-8.407			
10	9.401	20.762	-11.361			
11	14.48	26.282	-11.802			
12	22.328	24.524	-2.196			
13	15.298	18.644	-3.346			
14	15.073	17.51	-2.437			

The result suits my expectation. Because when I did the experiment, I found it was more difficult to perform the incongruent task.