

Test a Perceptual Phenomenon

February 9, 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 next section.

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

Independent variable: the condition, i.e. the congruent words condition and the incongruent words condition. Dependent variable: the time it takes to name the ink colors.

- (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 hypotheses(H_0): There is no significant difference in the time it takes to name the ink colors between the congruent words condition and the incongruent word condition. The alternative hypotheses(H_1): The time it takes to name the ink colors with the congruent condition is significantly different from it takes with the incongruent condition.

$$H_0 : t_1 = t_2$$

$$H_0 : t_1 \neq t_2$$

(t_1 is the time it takes to name the ink colors with the congruent condition, whereas t_2 is the time it takes to name the ink colors with the incongruent condition)

We should use dependent t-test for this task. Firstly, since we have a relatively small number of sample, 24, which is less than 30, and we do not know the parameter of the population, like the standard deviation, we should conduct a t-test, instead of a z-test. Secondly, we should choose dependent t-test instead of independent t-test, because in this test we have the same participants taking part in both of the tasks. In addition, since we assume the time it takes to name the ink colors is normally distributed in the population, and we did a random sampling from the population, all the requirements for the dependent t-test are satisfied. So in conclusion, we should conduct a dependent t-test for this task.

- (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 [1]: import pandas as pd
import matplotlib.pyplot as plt
data = pd.read_csv('stroopdata.csv')
data.describe()
```

```
Out[1]:
```

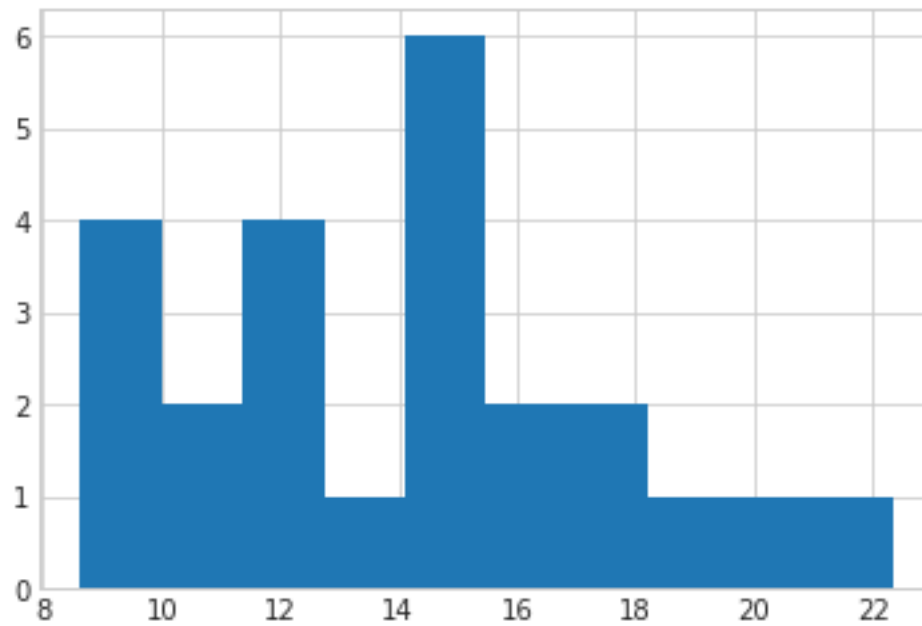
	Congruent	Incongruent
count	24.000000	24.000000
mean	14.051125	22.015917
std	3.559358	4.797057
min	8.630000	15.687000
25%	11.895250	18.716750
50%	14.356500	21.017500
75%	16.200750	24.051500
max	22.328000	35.255000

Under the congruent condition, the average time it takes the participants to name the ink color is 14.05 ($M = 14.05$), with the standard deviation of 3.56 ($SD = 3.56$). Under the incongruent condition, both of the statistics are larger, with the average time of 22.02 ($M = 22.02$) and the standard deviation of 4.80 ($SD = 4.80$). It seems that under the incongruent condition, participants tend to take longer time to finish the task and the variance between participants is larger too.

- (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.

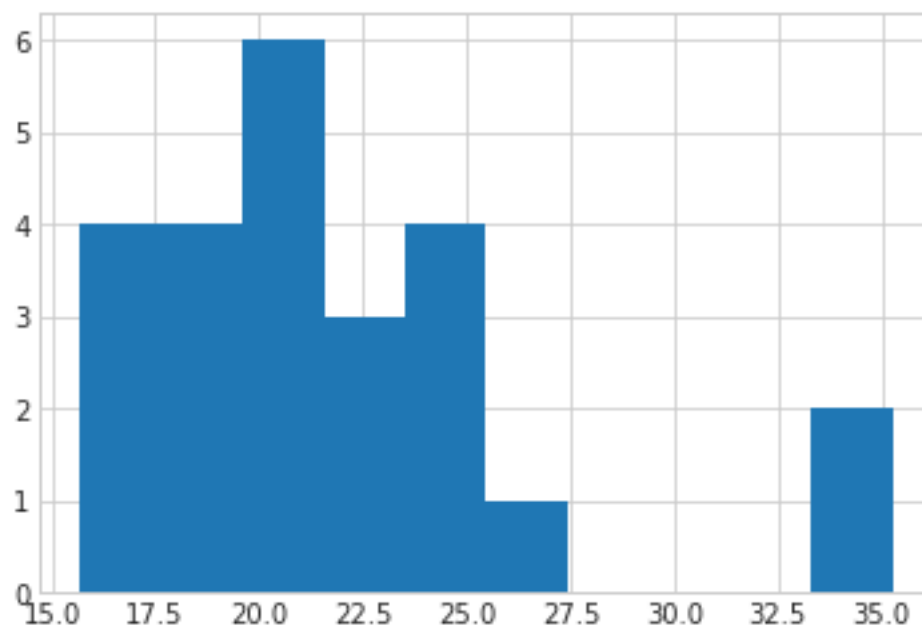
```
In [2]: %matplotlib inline
import seaborn as sns
sns.set_style('whitegrid')
plt.hist(data['Congruent'])
```

```
Out[2]: (array([ 4.,  2.,  4.,  1.,  6.,  2.,  2.,  1.,  1.,  1.]),
array([ 8.63,  9.9998, 11.3696, 12.7394, 14.1092, 15.479,
16.8488, 18.2186, 19.5884, 20.9582, 22.328 ]),
<a list of 10 Patch objects>)
```



```
In [3]: plt.hist(data['Incongruent'])
```

```
Out[3]: (array([ 4.,  4.,  6.,  3.,  4.,  1.,  0.,  0.,  0.,  2.]),
         array([ 15.687,  17.6438,  19.6006,  21.5574,  23.5142,  25.471 ,
                27.4278,  29.3846,  31.3414,  33.2982,  35.255 ]),
         <a list of 10 Patch objects>)
```



Both of the distribution of the congruent and incongruent condition are approximately normalized.

- (5) Now, perform the statistical test and report the results. What is the 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 scipy.stats as stats
        stats.ttest_rel(data['Congruent'],data['Incongruent'])
```

```
Out[4]: Ttest_relResult(statistic=-8.020706944109957, pvalue=4.1030005857111781e-08)
```

If we set the confidence level to be 95% ($\alpha=0.05$), according to the t-table, with the degrees of freedom 23, the critical statistic value is ± 2.069 . The calculated statistics is ± 8.02 , so we reject the null hypothesis. So in conclusion, the time it takes to name the ink colors with the congruent condition is significantly different from it takes with the incongruent condition. And it matches with my expectation.