

Test a Perceptual Phenomenon

March 5, 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?

The independent variable (x) whether or not the colors and the words are the same. The dependent variable (y) is 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.

H0: The mean time for the population to read back the incongruent and congruent data will be the same. (C = I) HA: The mean time for the population to read back the incongruent and congruent data will be different. (C != I)

I will be using a paired t test since the two variables are dependent, I am assuming a normal distribution, we do not know the population standard deviation, and I don't have 30+ samples to produce a reliable z score.

- (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 [2]: # Perform the analysis here
import pandas as pd
```

```
In [3]: df = pd.read_csv('stroopdata.csv')
```

```
In [4]: df.head()
```

```
Out[4]:
```

	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

```
In [5]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 24 entries, 0 to 23
Data columns (total 2 columns):
Congruent      24 non-null float64
Incongruent    24 non-null float64
dtypes: float64(2)
memory usage: 464.0 bytes
```

There are two columns and 24 non null values in each column

```
In [6]: df.describe()
```

```
Out[6]:
```

	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

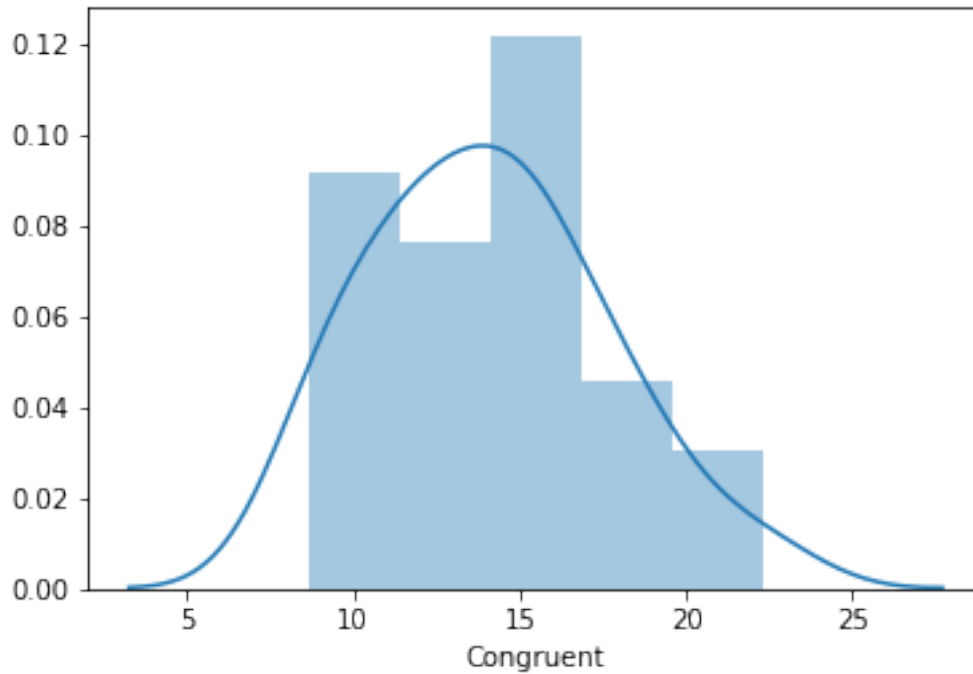
The median (50%) for congruent data is 14 while the mean for incongruent data is 21.

- (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 [7]: # Build the visualizations here
import math
import numpy as np
import seaborn as sns
from scipy import stats
get_ipython().magic('matplotlib inline')
```

```
In [8]: sns.distplot(df['Congruent'])
```

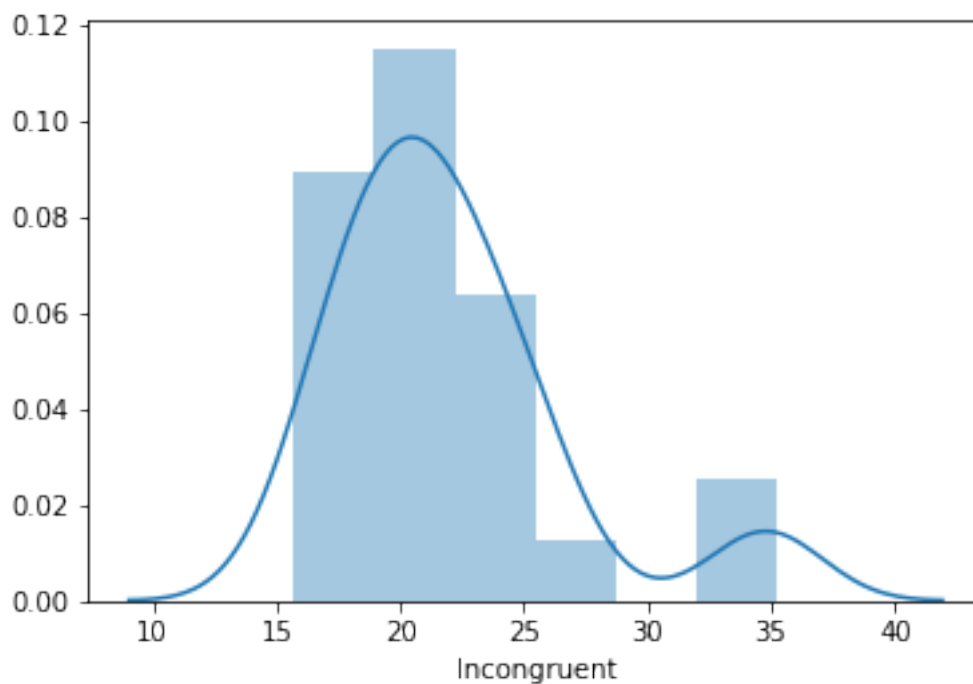
```
Out[8]: <matplotlib.axes._subplots.AxesSubplot at 0x7f30fc106160>
```



Shape: The shape is mostly normally distributed Center: We can visually see the median sits at around 14. Spread: We see that the data is spread between around 9 and 22

```
In [9]: sns.distplot(df['Incongruent'])
```

```
Out[9]: <matplotlib.axes._subplots.AxesSubplot at 0x7f30fc092668>
```



Shape: The data is right skewed. Center: We can visually see the median sits at around 21. Spread: We see that the data is spread between around 16 and 35 Based on the graphs it appears that it takes longer to read back the incongruent data and it may be responsible for the outliers.

- (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 [10]: # Perform the statistical test here
```

```
In [11]: df.head()
```

```
Out[11]:
```

	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

```
In [12]: stats.ttest_rel(df['Congruent'],df['Incongruent'])
```

```
#Resource: https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.ttest\_rel.h
```

```
Out[12]: Ttest_relResult(statistic=-8.020706944109957, pvalue=4.1030005857111781e-08)
```

```
In [16]: print (stats.t.ppf(1-0.025, 23))
```

```
#Resource: https://stackoverflow.com/questions/19339305/python-function-to-get-the-t-st
```

```
#Resource: https://statistics.laerd.com/statistical-guides/hypothesis-testing-3.php
```

```
#Resource: https://stats.idre.ucla.edu/other/mult-pkg/faq/general/faq-what-are-the-diff
```

```
2.06865761042
```

Since the Test statistic: 8.0207 is larger than the Critical statistic value of 2.069 and the p-value is basically 0 this indicates there is no significant difference between time used in congruent and incongruent data and thus we can reject the null hypothesis and accept the alternative hypothesis that it takes longer to read the incongruent data.

Confidence level is set at 95 percent (1-Significance Level) Significance Level is 0.05 (Most statisticians pick .05 or .01) p-value: 4.1030005857111781e-08 (Really small number)

Do you reject the null hypothesis or fail to reject it? There is a 5% or less chance that the difference in the congruent /noncongruent testing times is as different as observed given the null hypothesis is true. I therefore reject the null hypothesis and accept the alternative hypothesis that it takes longer to read the incongruent data.

Come to a conclusion in terms of the experiment task. Conclusion: It takes people longer to read the incongruent data than the congruent data.

Did the results match up with your expectations? It matched my expectations as I viewed the graphs of congruent and incongruent data I could see the additional time that it took for each observation to complete the task.