

Test a Perceptual Phenomenon - Final Project Submission

December 23, 2018

0.0.1 Analyzing the Stroop Effect

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

Definition:

Independent Variable or Response Variable or Outcome Variable - is the variable we are interested in predicting

Dependent Variable or Explanatory Variable or Predictor Variable - is the variable that predicts the response

In our case: we want to determine or predict if on average the time taken to name ink colors is affected by whether or not the individual(s) is presented with a congruent or incongruent word during the test, therefore we can say that - the independent Variable is - the congruency of the color and text - the dependent Variable is - duration or time to recognise the colors or text

(2) What is an appropriate set of hypotheses for this task? Specify your null and alternative hypotheses, and clearly define any notation used. Justify your choices.

In our case the appropriate alternate hypothesis or what we would like to test is whether, on average, the time taken by the population to name a congruent and incongruent word differs. That is :

H_0 : There is no significant difference in time for the population to state the colors of the words in a congruent or incongruent condition

H_1 : There is a significant difference in time for the population to state the colors of the words in a congruent or incongruent condition

Mathematically that is :

$$H_0 : \mu_{cong} = \mu_{incong}$$

$$H_1 : \mu_{cong} \neq \mu_{incong}$$

The above can re-ordered as :

$$H_0 : \mu_{cong} - \mu_{incong} = 0$$

$$H_1 : \mu_{cong} - \mu_{incong} \neq 0$$

(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 [11]: #load data and neccessary modules
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

```
import seaborn as sns
from scipy import stats

%matplotlib inline

df = pd.read_csv('data/stroopdata.csv')
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 [5]: df.mean() #on average, it take longer to name an incongruent word
```

```
Out[5]: Congruent      14.051125
Incongruent    22.015917
dtype: float64
```

```
In [7]: df.median() #at the 50% percentile - it still takes longer to respond to incongruent w
```

```
Out[7]: Congruent      14.3565
Incongruent    21.0175
dtype: float64
```

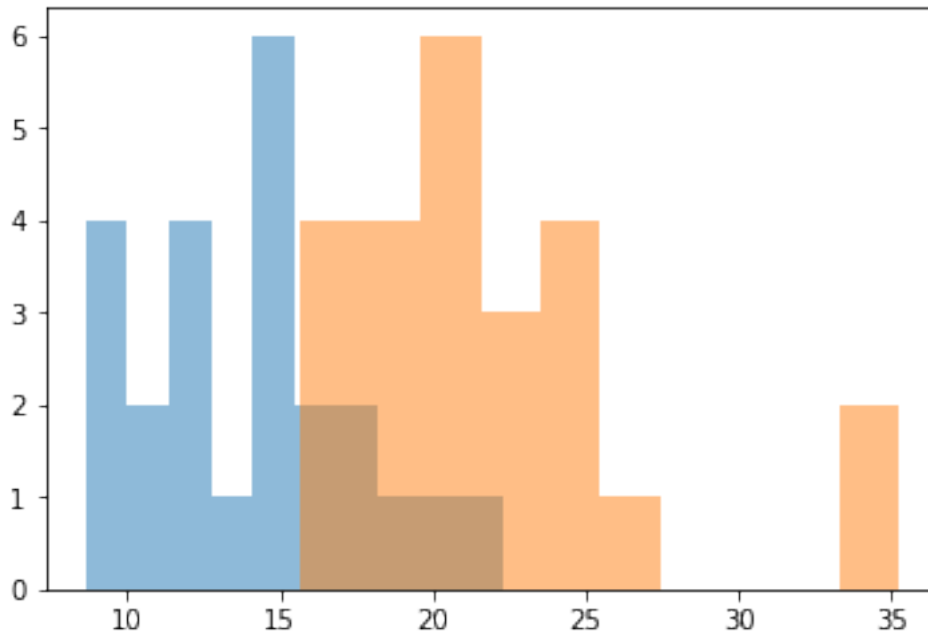
```
In [12]: #set ddof = 1 as the data is of a sample not a population
df.std(ddof=1) #seems the incongruent data is more spread
```

```
Out[12]: Congruent      3.559358
Incongruent    4.797057
dtype: float64
```

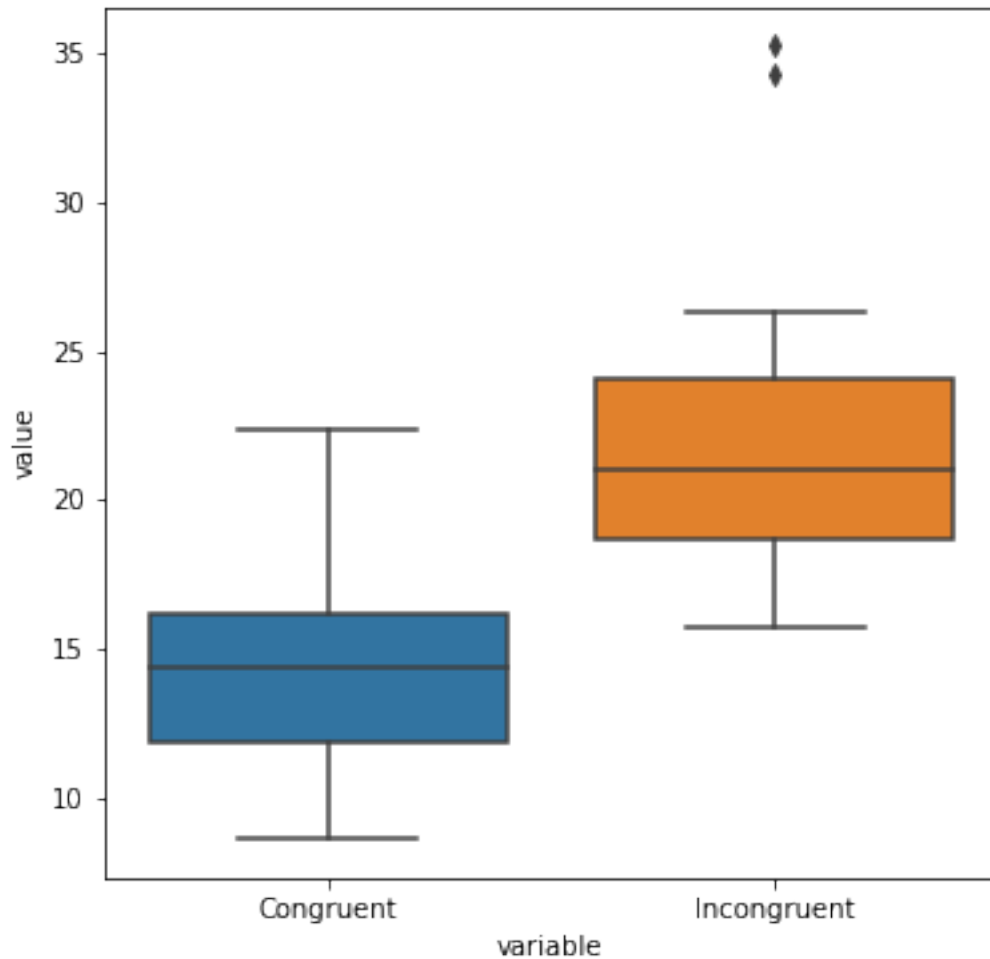
Descriptive Statistics Summary : On average - an individual takes longer (22.01) to name an incongruent word, compared to 14.05 on naming congruent words

(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 [8]: #create a histogram - to observe the shape of our data
plt.hist(df['Congruent'], alpha=0.5)
plt.hist(df['Incongruent'], alpha=0.5);
```



```
In [10]: #create a histogram and boxplots - to observe the shape of our data
#Boxplots multiple columns - http://bit.ly/2MPIxa8
#Resize Figure - http://bit.ly/2pjhYR6
fig, ax = plt.subplots()
# the size of A4 paper
fig.set_size_inches(6, 6)
sns.boxplot(x="variable", y="value", data=pd.melt(df), ax=ax);
```



Shape : Congruent Data and Incongruent Data: Congruent seems to be skewed to the right while the incongruent data is left-skewed with some outliers higher spectrum of the data

(5) Now, perform the statistical test and report your results. What is your confidence level or Type I error associated with your test? What is your conclusion regarding the hypotheses you set up? Did the results match up with your expectations? Hint : Think about what is being measured on each individual, and what statistic best captures how an individual reacts in each environment.

Selected Test Statistic : The test statistic that best captures how an individual reacts in each environment (that on congruent and incongruent words) is the **Dependent T-test for Paired samples**

Dependent T-test for Paired samples: is used when there is only one sample that has been tested twice (that is we have repeated measures) or two samples that have been matched or paired

Assumptions This test assumes : 1. The differences are of the measured variable, in our case we are using the differences in the dependent variable that is duration taken to identify colors and

text 2. Sampling is random and the pairs of observations are independent 3. The difference in means is normally distributed

```
In [3]: #perform statistical analysis - http://bit.ly/2pl6dcW - Scipy ttest_rel
stats.ttest_rel(df['Congruent'], df['Incongruent'])
```

```
Out[3]: Ttest_relResult(statistic=-8.020706944109957, pvalue=4.103000585711178e-08)
```

Inferential Statistics Conclusions Our Type I Error threshold is $\alpha = 0.05$ (or 95% confidence interval) - we observe a p-value less than our α therefore we reject our null hypothesis and conclude that there is statistical evidence that on average the response times when identifying the congruent and incongruent words significantly differs.

This is in line with what we expected, that is the individuals would take longer to name the incongruent words than the congruent words.

(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!

What is responsible for above effects : The above effects of individuals taking a longer on naming incongruent words (compared to congruent words) can be attributed to the fact that humans are involuntary readers - we see words before anything else. Therefore it would be expected identifying colors on congruent words would require much less effort than incongruent words.

Alternative Experiments with similar results : In marketing, advertisers will go words that make their goods and services seem better to the target niche using words like 'the better option', 'simple', 'reliable' or 'fast' than their rivals. Customers find themselves going for such goods and services even when they may not be that different from what their competitor are offering.