

Test a Perceptual Phenomenon

March 16, 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 final section of this lesson.

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

0.0.2 --answer--

In an experiment, researchers manipulate the independent variables, measure changes in the dependent variable and seek to control lurking variables. In this case, the independent variable is the incongruent/ the congruent words condition and the dependent variable is "Time it takes to name the ink color".

- (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.

0.0.3 --answer--

The null hypotheses is that there is no statistically significant difference between the sample mean of the time it takes for the participant to do the test given the congruent words condition and that of the incongruent words condition: $H_0 : \mu = \mu_I$

So, the alternative hypotheses could be any of the following set:

- 1- The mean of the time it takes for the participant to do the test given the **congruent** words condition is longer than that of the **incongruent** words condition:

$$H_a : \mu > \mu_I$$

- 2- The mean of the time it takes for the participant to do the test given the **incongruent** words condition is longer than that of the **congruent** words condition:

$$H_a : \mu < \mu_I$$

- 3- The mean of the time it takes for the participant to do the test given the **incongruent** words condition is significantly different than that of the **congruent** words condition:

$$H_a : \mu \neq \mu_I$$

In this experiment, we use this last alternative hypothesis ($H_a : \mu \neq \mu_I$), so we do not care about the directionality of the alternative hypothesis.

- (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 [39]: # Perform the analysis here
import pandas as pd
df = pd.read_csv('stroopdata.csv')
df.head()
```

```
Out[39]:
```

	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 [50]: df.median()
```

```
Out[50]:
```

Congruent	14.3565
Incongruent	21.0175

dtype: float64

```
In [38]: df.describe()
```

```
Out[38]:
```

	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

0.0.4 --answer--

the above table has provided some very useful stats about the dataset for both Congruent and Incongruent data sets

- (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.

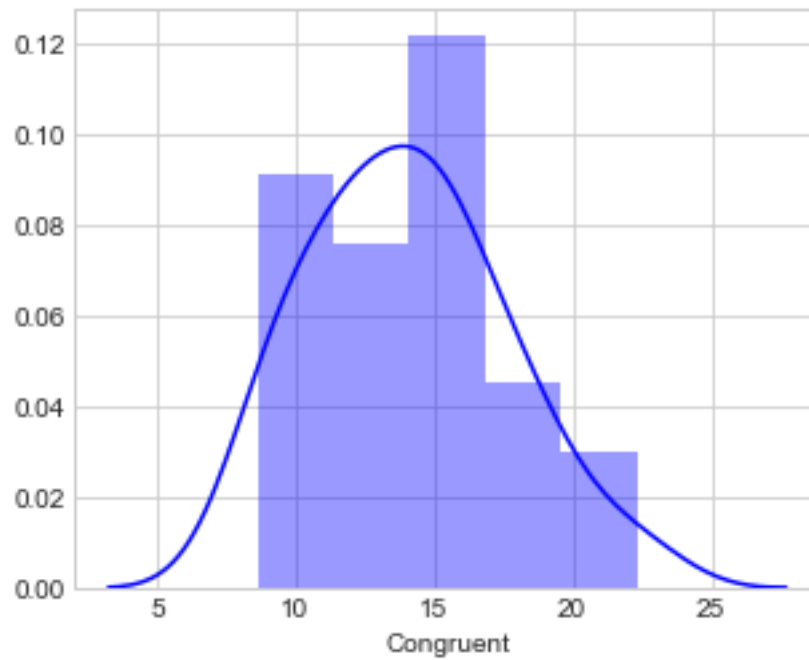
```
In [37]: # Build the visualizations here
import matplotlib.pyplot as plt
from matplotlib import rcParams
import matplotlib as mpl
import seaborn as sb
import numpy as np
```

```
%matplotlib inline
rcParams['figure.figsize'] = 5, 4
sb.set_style('whitegrid')

congruent_data = df.Congruent
incongruent_data = df.Incongruent

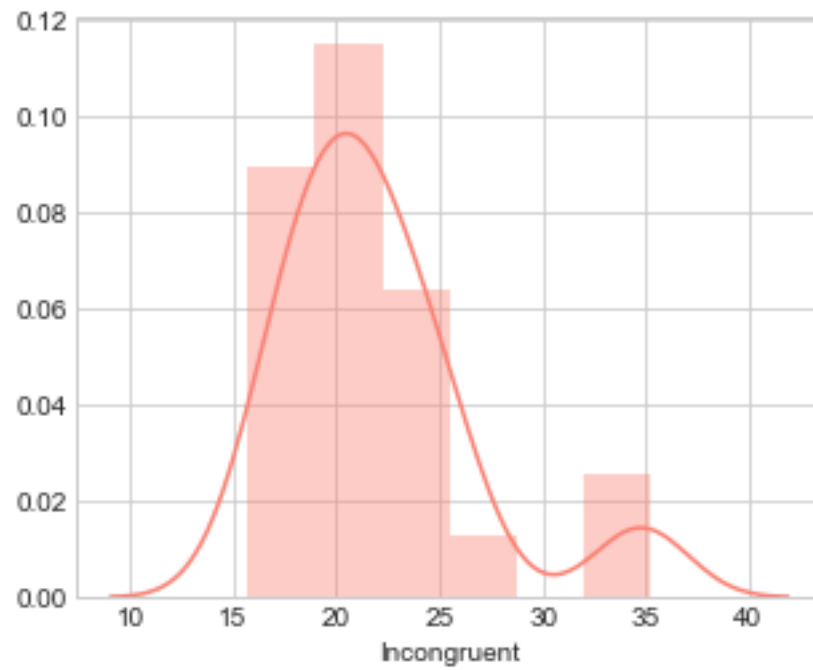
sb.distplot(congruent_data, color='blue')
```

Out[37]: <matplotlib.axes._subplots.AxesSubplot at 0x2399912d630>



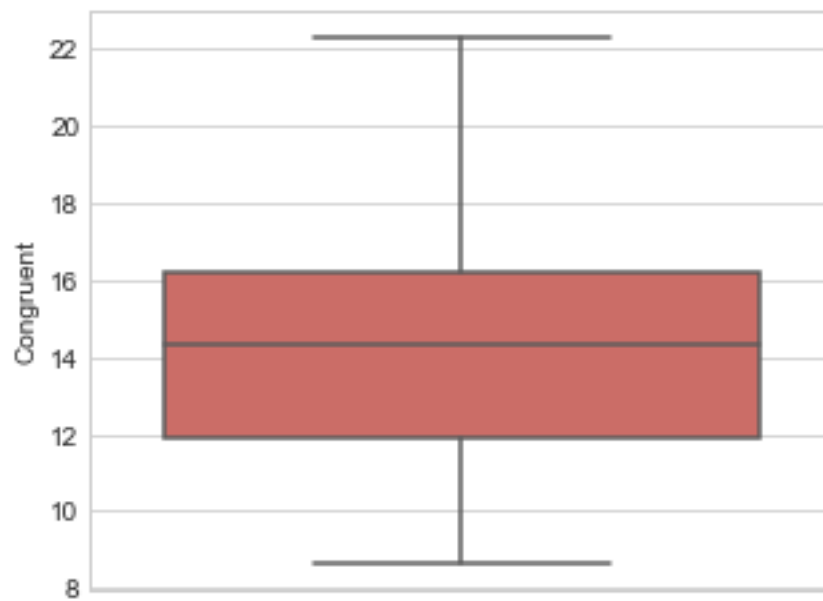
```
In [36]: sb.distplot(incongruent_data, color='salmon')
```

Out[36]: <matplotlib.axes._subplots.AxesSubplot at 0x2399ad6bcc0>



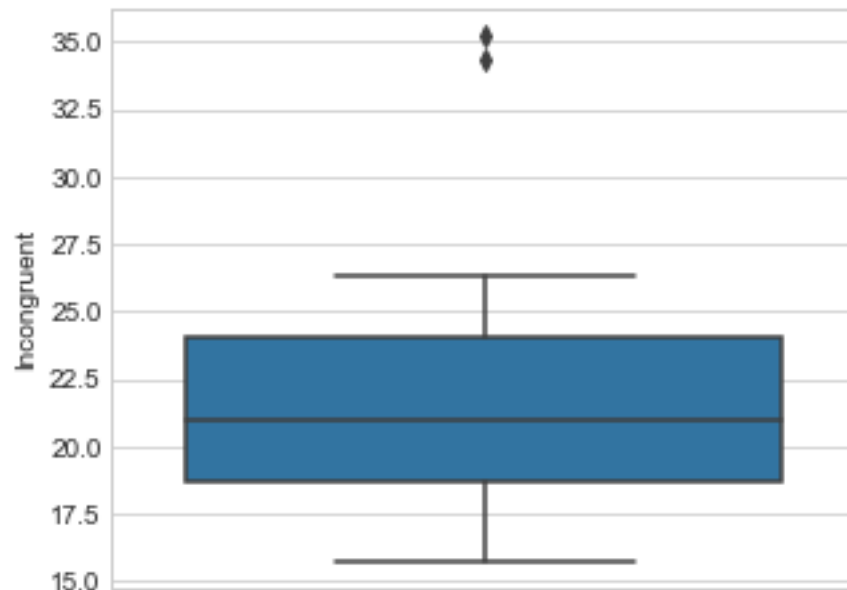
```
In [42]: sb.boxplot(y='Congruent', data=df, palette='hls')
```

```
Out[42]: <matplotlib.axes._subplots.AxesSubplot at 0x2399b05c4a8>
```



```
In [44]: sb.boxplot(y='Incongruent', data=df)
```

```
Out[44]: <matplotlib.axes._subplots.AxesSubplot at 0x2399af1e860>
```



0.0.5 --answer--

As we can see from the plots above, the two samples have rather different distribution (Shown by Histogram plots). Incongruent dataset has a median value of around 21 but the congruent set has a median of approximately 14 which is apparent in the Box Plots. This shows that on average it took longer for the participants to complete the test in case of the Incongruent condition compared to that of the congruent condition.

- (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.

```
In [49]: # Perform the statistical test here
```

```
import scipy.stats as stats
```

```
statistic, pvalue = stats.ttest_ind(congruent_data, incongruent_data)
```

```
print('pvalue = ', pvalue)
```

```
pvalue = 4.5949489622951854e-08
```

0.0.6 --answer--

We use the t-test for two independent samples from a population. The test measures whether the average (expected) value differs significantly across samples. If we observe a large p-value, for instance larger than 0.05, then we retain the null hypothesis of identical average scores.

In this case, we have observed a very small p-value indicating that there is a statistically significant difference between the two sample means, so we reject the null hypothesis of equal averages.

- (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!

--write answer here--