

Test a Perceptual Phenomenon

June 21, 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, upload that PDF/HTML into the workspace here (click on the orange Jupyter icon in the upper left then Upload), then use the Submit Project button at the bottom of this page. This will create a zip file containing both this .ipynb doc and the PDF/HTML doc that will be submitted for your project.

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

–write answer here–

Independent variable: Inside the congruent/incongruent group, each participant's used time.
Dependent variable: For each participant, his/her time usage for congruent and incongruent test.

(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.

–write answer here–

(a) Null hypothesis: There is no significant difference in time for the population to state the colors of the words in a congruent or incongruent condition.

$H_0: \text{congruent} = \text{incongruent}$

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

$H_1: \text{congruent} \neq \text{incongruent}$

(b) For this dataset, I choose to use T-test.

First, we assume the two population (response times for congruent and incongruent words for any person) and samples are normally distributed.

Then in this case, we try to use two groups of related samples to compare and guess if their population means are equal or not. In the meanwhile, the variances of two populations are unknown, and the sample size is pretty small, so we cannot use z-test. So I chose to use t-test, because t-test can identify the difference of means of two population with small samples and unknown variance. Also, the samples are paired (for the same participant with two observation in different conditions), so I chose to use paired t-test.

(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 [22]: # Perform the analysis here
import seaborn as sns
import pandas as pd
import matplotlib.pyplot as plt
import scipy.stats as st

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

```
Out[22]:
```

	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

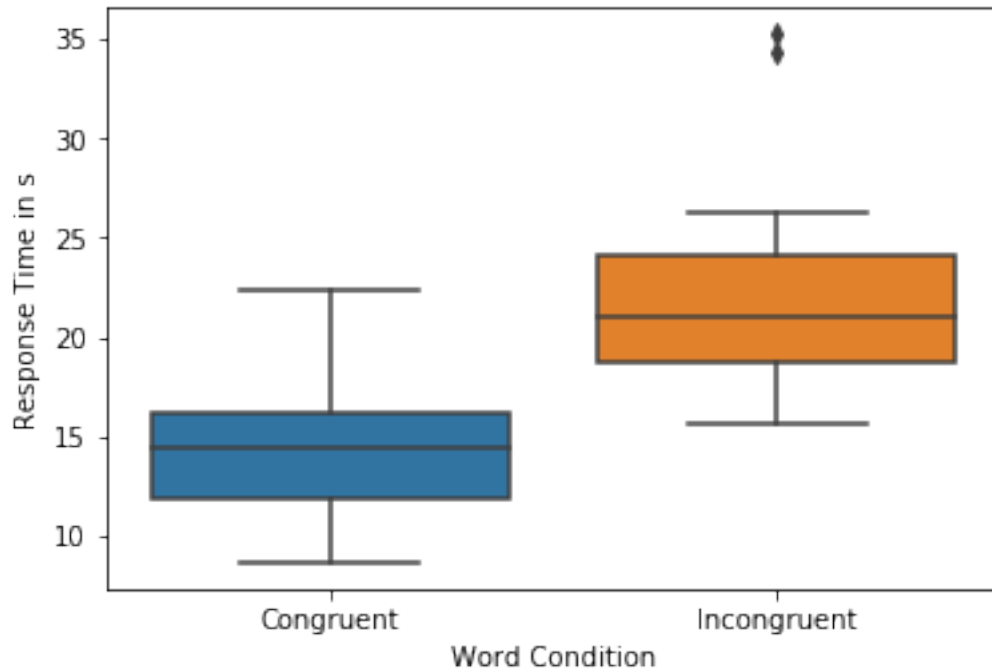
–write answer here–

- (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 [23]: cong = df["Congruent"]
incong = df["Incongruent"]
fig, ax = plt.subplots()
ax = sns.boxplot(data=pd.DataFrame({"Congruent":cong,
                                   "Incongruent":incong}))

plt.xlabel('Word Condition')
plt.ylabel('Response Time in s')
```

```
Out[23]: Text(0,0.5,'Response Time in s')
```



–write answer here– As the graph showing, the Response Time of Incongruent group is significantly higher than congruent group, so it is very likely that there is a significant difference in these two groups' means.

- (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 [24]: # Perform the statistical test here
         [t,p] = st.ttest_rel(cong,incong)
         print(t)
         print(p)
```

```
-8.02070694411
4.10300058571e-08
```

–write answer here–

I choose confidence level as 95%, so the significance level is 0.05.

t-statistic = -8.02070694411

p-value = 4.10300058571e-08

Conclusion: Because the p-value is less than 0.05, we can reject the null hypothesis, then we accept the alternative hypothesis, that is there is a significant difference in these two groups' means, which means the unmatched color ink("incongruent") does make participants take more time to recognize color 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!

–write answer here– Because the color of ink and the meaning of words are different and contradictory against each other, and this takes us longer time to respond. Similar task: try to recognize each words' ink color, with the word's meaning is unmatch.