

P1- Testing a Perceptual Phenomenon

CODE ▼

Ryan D. Watts

- **University** - Western Governors University
- **Course** - C749 - Introduction to Data Science
- **Publish Date** - 09/05/2017

Overview

The stroop task consists of completing a test by which contains a list of words displayed in an alternative color of ink or background that is different from the printed word. The purpose of this task/test is to analyze the length of time it takes to identify the printed word within a given word set. The test generally consists of two word conditions *congruent* vs. *incongruent*.

This document contains an analysis of the stroop effect by measuring sample data containing incongruent vs. congruent time results.

The project contains a set of **Questions for Investigation**, for consistency purposes the list of questions will be provided below and can be identified with a header following this pattern `Q# - Question`.

Questions for Investigation

1. What is our independent variable? What is our dependent variable?
2. What is an appropriate set of hypotheses for this task? What kind of statistical test do you expect to perform? Justify your choices.
3. Report some descriptive statistic regarding this dataset. Include at least one measure of central tendency and at least one measure of variability.
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.
5. Now, perform the statistical test and report your results. What is your 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.

Q1: What is our independent variable? What is our dependent variable?

The given data set `stroopdata.csv` contains both an independent and dependent variable. Although at first glance the independent variable may not appear directly, we can determine based on contextual information that the **independent variable** is the the word list which was used to generate the sample data set.

The **dependent variable** within the given data set is the response time in seconds from all participants (24).

Q2: What is an appropriate set of hypotheses for this task? What kind of statistical test do you expect to perform?

I have elected to use the `two-tailed t-test` for my hypothesis.

Two-Tailed T-Test Reason for Selection

- Due to lack of randomization within the given dataset and they were apart of a given set the two-tailed t-test is appropriate
- The ability to compare the means of the dataset for the pre and post test validates the benefit of this test selection

Hypothesis

As we prepare to analyze the hypothesis, there are two variables that will be used. The mean of incongruent words within the dataset will be denoted as \bar{x}_i ¹. While the mean of congruent words within the dataset is denoted as \bar{x}_c ².

- Alternative Hypothesis:

$$H_A = \text{UnkownCharacter}_i - \text{UnkownCharacter}_c \neq 0$$

Because there isn't a relative difference regarding the response time in the congruent vs. incongruent words the null hypothesis can be determined by the equation provided above.

- Null Hypothesis:

$$H_O = \text{UnkownCharactor}_i - \text{UnkownCharacter}_c = 0$$

The response time in regards to the incongruent and congruent words are significantly different when analyzing each variable. The response time's speed has no regards to the response time provided during the pre-test period.

Q3 Report some descriptive statistic regarding this dataset. Include at least one measure of central tendency and at least one measure of variability.

In order to accurately report a descriptive statistic we must first define variables that will be used for reporting.

Define Variables for Descriptive Analysis

HIDE

```
# load sample data and assign congruent incongruent
sampledata <- read.csv("stroopdata.csv");
congruent = sampledata$Congruent;
incongruent = sampledata$Incongruent;
# define central/variability tendencies
congruent_mean = round(mean(congruent), 2);
congruent_median = round(median(congruent), 2);
congruent_sd = round(sd(congruent), 2);
incongruent_mean = round(mean(incongruent), 2);
incongruent_median = round(median(incongruent), 2);
incongruent_sd = round(sd(incongruent), 2);
sampledata_summary = summary(sampledata);
# altogether , print out each variable in it's context
cat(str_interp("
Congruent Mean: ${congruent_mean}
Congruent Median: ${congruent_median}
Congruent Standard Deviation: ${congruent_sd}
Incongruent Mean: ${incongruent_mean}
Incongruent Median: ${incongruent_median}
Incongruent Standard Deviation: ${incongruent_sd}
"))
```

```
Congruent Mean: 14.05
Congruent Median: 14.36
Congruent Standard Deviation: 3.56

Incongruent Mean: 22.02
Incongruent Median: 21.02
Incongruent Standard Deviation: 4.8
```

HIDE

```
summary(congruent)
```

| Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
|------|---------|--------|-------|---------|-------|
| 8.63 | 11.90 | 14.36 | 14.05 | 16.20 | 22.33 |

HIDE

```
summary(incongruent)
```

| Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
|-------|---------|--------|-------|---------|-------|
| 15.69 | 18.72 | 21.02 | 22.02 | 24.05 | 35.26 |

HIDE

```
# show the data in a table for quick analysis
```

```
DT::datatable(data=sampladata)
```

Show 10 entries

Search:

| | Congruent | Incongruent |
|----|-----------|-------------|
| 1 | 12.079 | 19.278 |
| 2 | 16.791 | 18.741 |
| 3 | 9.564 | 21.214 |
| 4 | 8.63 | 15.687 |
| 5 | 14.669 | 22.803 |
| 6 | 12.238 | 20.878 |
| 7 | 14.692 | 24.572 |
| 8 | 8.987 | 17.394 |
| 9 | 9.401 | 20.762 |
| 10 | 14.48 | 26.282 |

Showing 1 to 10 of 24 entries

Previous

1

2

3

Next

Q4: 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.

HIDE

```
#plot to show the distribution of data and comparison between congruent vs. incongruent.
```

```
p1 <- plot_ly(sampladata, x = ~congruent, type = "box", name = "Congruent") %>%
```

```
  add_trace(name = "Incongruent", x = ~incongruent, type = "box") %>%
```

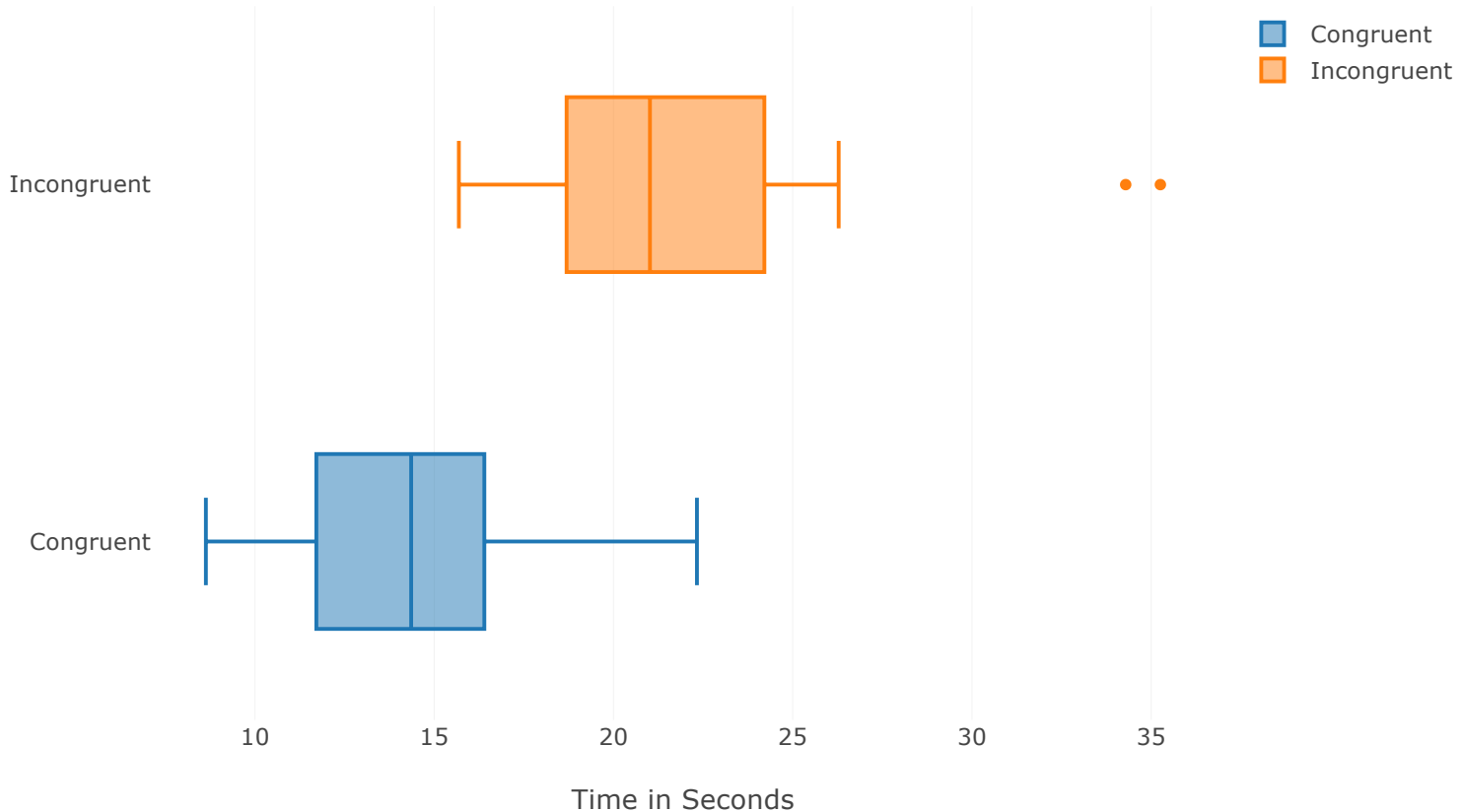
```
  layout(xaxis = list(title = "Time in Seconds"), width = 800, autosize = F, margin = list( l = 10  
0, b = 50))
```

Specifying width/height in layout() is now deprecated.

Please specify in ggplotly() or plot_ly()

HIDE

p1



Observing the Plot

- While analyzing the box plots we can determine that there are two distinct outliers within the Incongruent dataset.
- We can also determine that the Higher Third Quartile of the Congruent dataset is consistent with the Median Quartile of the Incongruent data.

Q5: Now, perform the statistical test and report your results. What is your 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.

variables

$$\alpha = 0.1$$

$$Df = 24$$

t-statistic

$$t - critical = +1.714 - 1.714$$

In order to properly conduct a statistical analysis of the data, we must define variables that will be used to calculate our findings. The variables defined below will be used to determine the conclusion on the next page.

```
#prep data for statistical test
est <- round((incongruent_mean - congruent_mean), 2);
#create a new row on data table for the difference between incongruent and congruent
sampledata["Difference"] = sampledata["Incongruent"] - sampledata["Congruent"];
#create the standard deviation difference of congruent and incongruent data
sddiff <- round(sd(sampledata$Difference) ,3);
#determine t-statistic using the formula given above.
tstat <- round(est/(sddiff/(sqrt(nrow(sampledata)))), 3);
```

Conclusion

During the statistical analysis I started by evaluating the t-critical which was determined to be 8.026. If we take another look at the formula defined above we would expect a the value to be < 1.714 in order for our **Null Hypothesis** to be accepted. According to the formula our results are $t - statistic = 8.026 > 1.714$. Due to these findings we have no choice but to reject our null hypothesis. If we recall the null hypothesis stated that there were no significant differences in the response time between congruent and incongruent data.

1. mean of incongruent - \bar{y} ↩

2. mean of congruent - \bar{c} ↩