

# Udacity Project – Stroop Effect

## Statistics: The Science of Decisions Project Instructions

### Background Information

In a Stroop task, participants are presented with a list of words, with each word displayed in a color of ink. The participant's task is to say out loud the *color of the ink* in which the word is printed. The task has two conditions: a congruent words condition, and an incongruent words condition. In the *congruent words* condition, the words being displayed are color words whose names match the colors in which they are printed: for example, **RED**, **BLUE**. In the *incongruent words* condition, the words displayed are color words whose names do not match the colors in which they are printed: for example, **PURPLE**, **ORANGE**. In each case, we measure the time it takes to name the ink colors in equally-sized lists. Each participant will go through and record a time from each condition.

### Questions for Investigation

As a general note, be sure to keep a record of any resources that you use or refer to in the creation of your project. You will need to report your sources as part of the project submission.

1. What is our independent variable? What is our dependent variable?

**Answer:**

- Independent variable = Colored congruent and incongruent words
- Dependent Variable = Response Time

2. What is an appropriate set of hypotheses for this task? What kind of statistical test do you expect to perform? Justify your choices.

**Answer:**

**(2a) Hypothesis:**

- a. Null Hypothesis ( $H_0$ ) – There is no difference in the response time for congruent and incongruent scenarios

$H_0 = \mu_C = \mu_{IC}$  where  $\mu_C$  Congruent population mean,  $\mu_{IC}$  Incongruent population mean

- b. Alternate Hypothesis ( $H_a$ ) – There is significant difference in the response time for congruent and incongruent scenarios

$H_0 = \mu_C \neq \mu_{IC}$  where  $\mu_C$  Congruent population mean,  $\mu_{IC}$  Incongruent population mean

**(2b) Statistical Test and Justification:**

We will be conducting a **t-test** instead of z- test as the number of samples provided are less than 30 and the standard deviation of the population is unknown.

The experiment was conducted on the same user groups for two different use cases [Congruent/Incongruent Words]. This data set constitute a pre-and post-scenario];

As a result, we will be conducting a **Paired-Sample T test**.

- Population Standard Deviation ( $\sigma$ ) = Unknown
- Samples Size < 30
- Paired Sample/ Pre/Post 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.

**Answer:**

We conducted a Paired Sample Two Tailed t-test with alpha = 0.05, 95% CI

Xc - Congruent Samples

Xi – Incongruent Samples

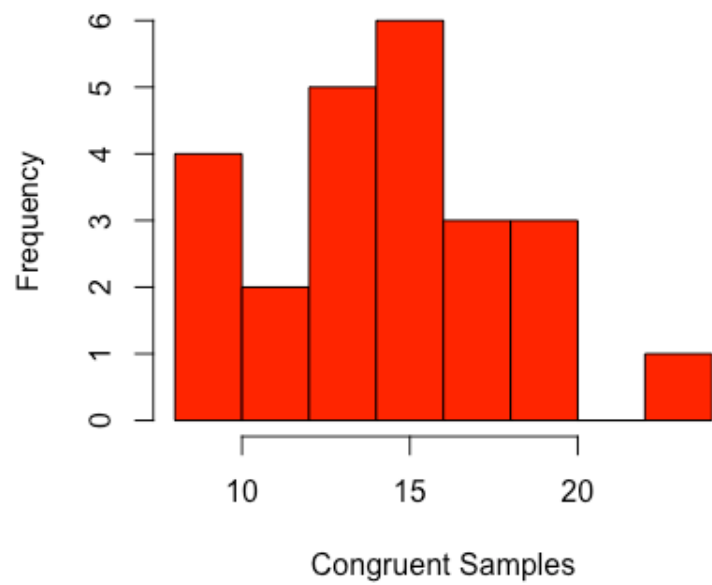
Please find below the values calculated for the t- test

- Mean of the Difference ( $\bar{x}$ ) =  $\overline{Xc - Xi} = -7.964791667$
- Standard Deviation (S) of the difference = 4.86482691
- T statistics = -8.020706944
- T Critical = +- 2.069 [from t- Table]
- Cohen's D = -1.637219949
- Margin of Error = -1.727747086
- R2 (coefficient of determination) = 10.84245719%

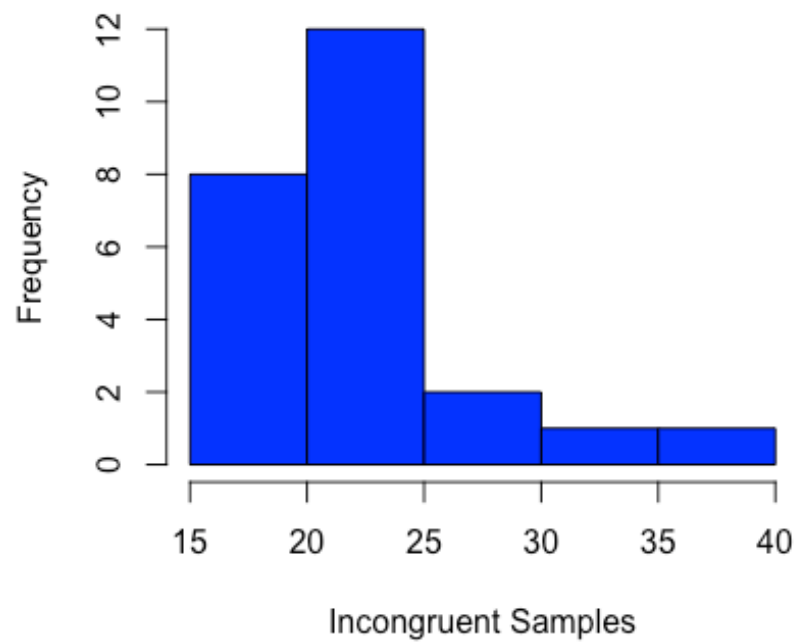
4. Provide one or two visualizations that show the distribution of the sample data. Write one or two sentences noting what you observe about the plot or plots.

**Answers:** Graphs are plotted using R-Studio for the data sample provided

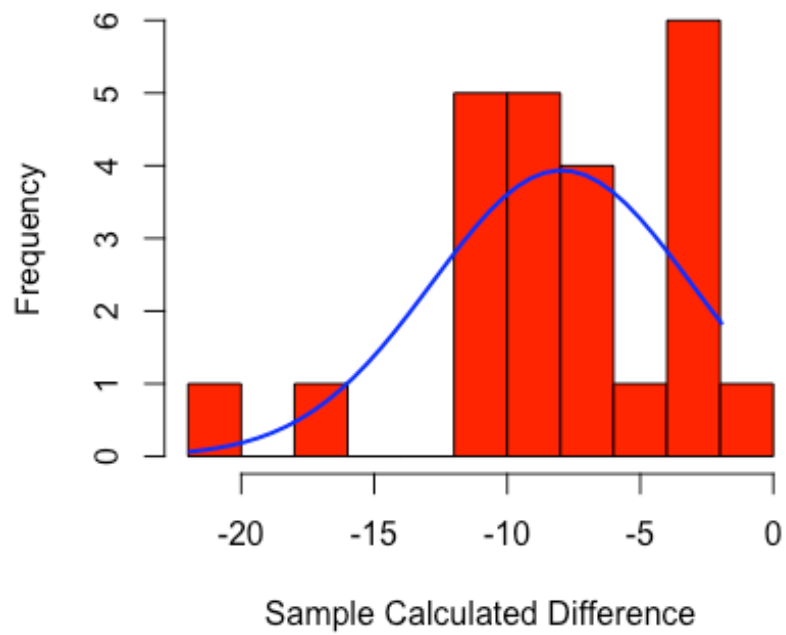
**Congruent Data Set Histogram**



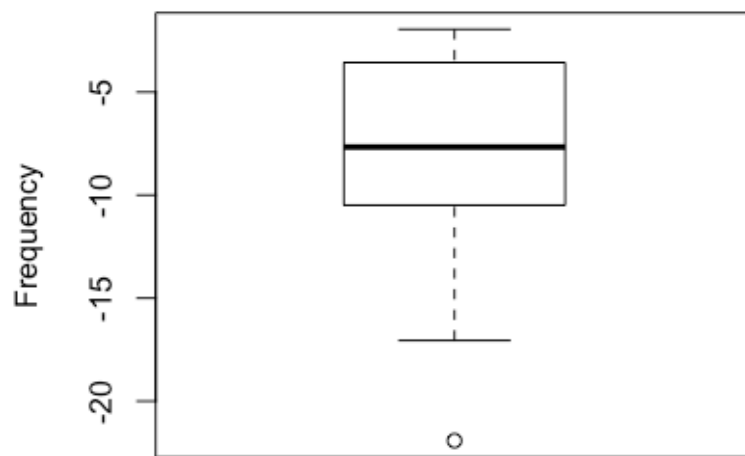
**Incongruent Data Set Histogram**



**Derived Data Set Histogram**



**Derived Data Set**



**Difference of Sample Data**

The sample data set for the congruent and incongruent are right or positive skewed distribution. The derived data set distribution is left or negatively skewed and so the box plot.

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? Conclude in terms of the experiment task. Did the results match up with your expectations?

**Answer:**

The t – test values are calculated by using Spreadsheet and R-Studio. P value using GraphPad

**Hypothesis:**

- a. Null Hypothesis (H<sub>0</sub>) – There is no difference in the response time for congruent and incongruent scenarios

$H_0 = \mu_C = \mu_{IC}$  where  $\mu_C$  Congruent population mean,  $\mu_{IC}$  Incongruent population mean

- b. Alternate Hypothesis (H<sub>a</sub>) – There is significant difference in the response time for congruent and incongruent scenarios

$H_0 = \mu_C \neq \mu_{IC}$  where  $\mu_C$  Congruent population mean,  $\mu_{IC}$  Incongruent population mean

**Statistical Test:** Paired Sample Two Tailed t-test with alpha = 0.05, 95% CI

**R-Studio Results:**

t = -8.0207, df = 23, p-value = 4.103e-08

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval: -10.019028 -5.910555

**P Value Results from GraphPad:**

t=-8.0207 DF=23. The two-tailed P value is less than 0.0001. By conventional criteria, this difference is extremely statistically significant.

Also as the mean difference is –VE I calculated the Paired One Tailed t-test with alpha = 0.05, 95% CI as well. Results below

**R-Studio Results – One Tail Test:**

t = -8.0207, df = 23, p-value = 2.052e-08

alternative hypothesis: true difference in means is less than 0

95 percent confidence interval:

-Inf -6.262868

**Conclusion:** The P value is far less than 0.05. The Tstatistic < Tcritical for 95% confidence levels. The difference in the mean is significant.

Therefore, we can confidently say there is significant difference in the time taken to read the congruent and incongruent words there by rejecting the null hypothesis (H<sub>0</sub>).

Also, based on the  $r^2$  results around 11% of the change is attributed by the congruent and incongruent words.

Based on the CI values there is 6 ~ 10 sec delay between reading the congruent and incongruent 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!

**Answer:**

The interpretation of words and colors by human brain and the effectiveness of brain recognizing the words and the colors are the factors responsible for the effects observed.

Similar Experiments:

- a. Reading a sentence or a word backwards
- b. Identifying words that are palindrome in a paragraph

**References:**

- a. <https://www.r-bloggers.com/>
- b. <http://www.graphpad.com/quickcalcs/>
- c. R- Studio
- d. Excel Spreadsheet
- e. Udacity Class Notes