Udacity Data Analyst Nanodegree
Project 1: Test a Perceptual Phenomenon

Questions for Investigation

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

- The independent variable of the Stroop task is whether the word set shown are congruent with the ink colors or if they are incongruent with the ink colors.
- The dependent variable of the Stroop task is the time it takes an individual to name the ink colors.

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

- The kind of statistical test that will be acceptable for this is task is the **two-tailed sample t-test**. The reasoning behind this decision is listed below.
    - First, we are comparing the average difference between the response time amongst the same 24 individuals in two different conditions. (congruent words and incongruent words)
    - With smaller sample sizes typically under 30, t-tests are a common choice. T-tests are common amongst smaller sample sizes, it is important to make the assumption that the mean of the overall population follows a normal distribution.

- **Hypotheses**

$$H_0: \mu_{difference} = 0$$
$$H_a: \mu_{difference} \neq 0$$

    - Null Hypothesis known as H_0
        - The null hypothesis states that there is no statistically significant difference amongst the mean time of both population conditions. (congruent and incongruent)

    - Alternative Hypothesis known as H_a
        - The alternative hypothesis assumes that there is a statistically significant difference between the mean times of both populations' conditions. (congruent and incongruent)

- $\mu_{difference}$ is the difference between the population mean time of the congruent data group and the incongruent data group.

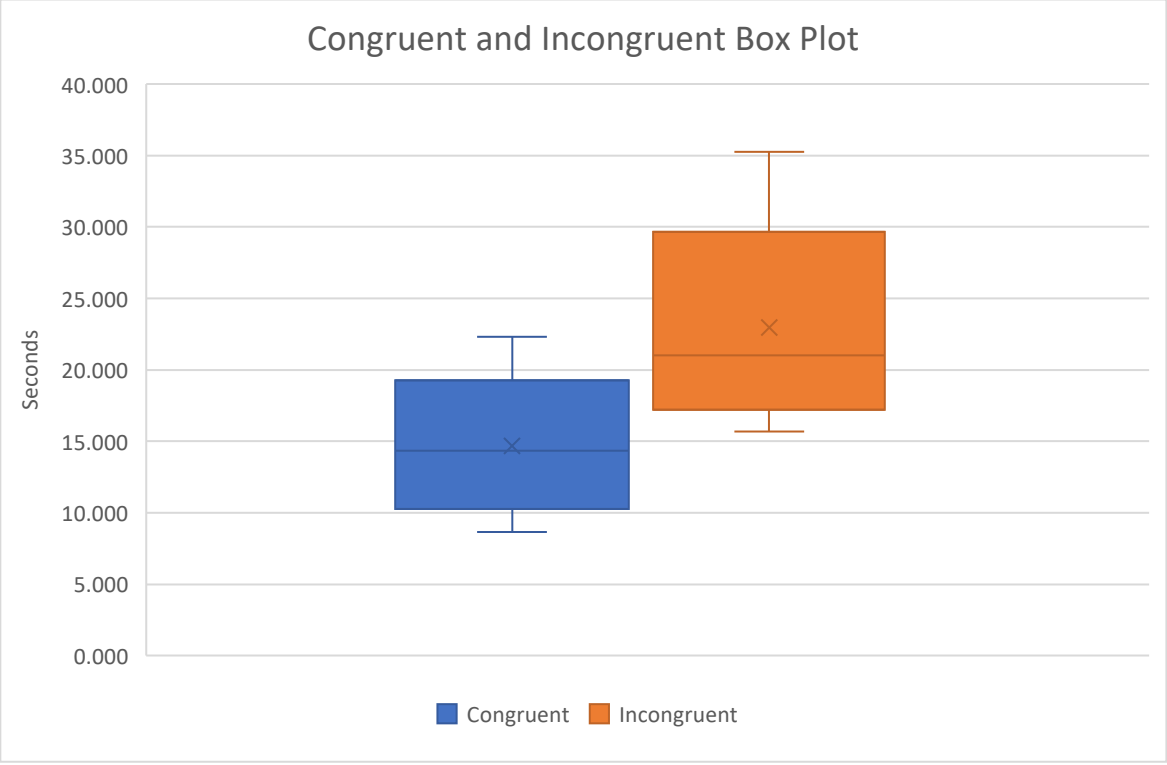
3. Report some descriptive statistics regarding this dataset. Include at least one measure of central tendency and at least one measure of variability.

Descriptive Statistics Summary

	Congruent	Incongruent
Sample Size	24.000	24.000
Mean	14.051	22.016
Standard Deviation	3.559	4.797
Minimum Value	8.630	15.687
25th Percentile	11.895	18.717
Median	14.357	21.018
75th Percentile	16.201	24.052
Maximum Value	22.328	35.255

The mean of congruent group is 14.051 seconds while the mean of the incongruent group is 22.016 seconds. The standard deviation of the congruent group is 3.559 seconds, and the standard deviation of the incongruent group is 4.797.

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.



It is evident the range between the minimum value and maximum value of the incongruent group is larger than that of the congruent group. This due to the fact that there are a few outliers amongst the group which extends the range.

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?

- Degree of freedom equation is df = sample size number – number of parameters calculated
    - Df = 24 – 1
    - Df = 23
- The confidence level will be set at **95%**.
- The p value is 0.05.
- The critical statistic value is - 2.069 and +2.069.
- The t statistic value is -8.021.
- We will reject the null hypothesis due to the fact that there is a significant difference between the means of the congruent and incongruent group. The values within the 95% confidence interval are -10.02 and -5.91, this was calculated by taking the t statistic value and subtracting and adding the t-critical value of -2.069 and +2.069 respectively. All of the information points to in the incongruent group the response time was worse than that of the congruent group.

Resources

- https://faculty.washington.edu/chudler/java/ready.html
- https://www.statisticshowto.com/how-to-do-a-t-test-in-excel/