Test a Perceptual Phenomenon

March 19, 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?

The independent variable consist of the two sets of words: the one with the congruent words condition and the one with the incongruent words condition.

The dependent variable consist of the times it takes the reader to go through all the words of each set.

(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.

Null: It takes the same average time to read the incongruent list of words than the average time it takes to read the congruent list of words.

Alternative: The average time it takes to read the incongruent list of words is different than the time it takes to read the congruent list of words.

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

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

μ symbol is used to represent the mean. H_0 represents the null hypothesis. H_1 represents the alternative.

(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'.

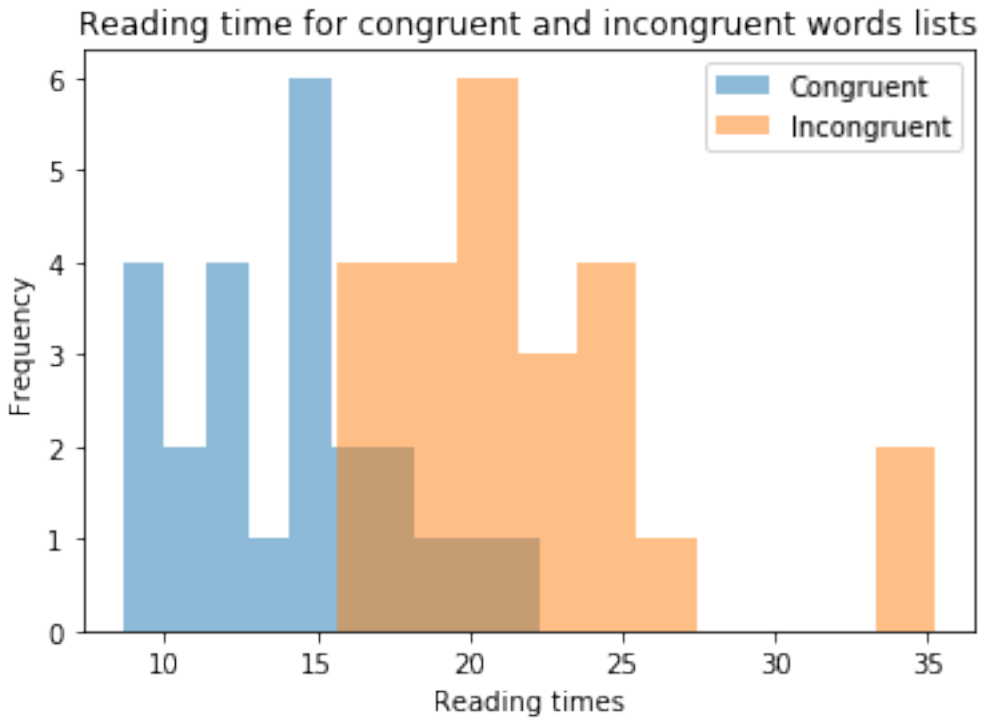
```
In [1]: import pandas as pd
        import matplotlib.pyplot as plt
        import numpy as np
        from scipy import stats

        df = pd.read_csv('stroopdata.csv')

        df.describe()
```

```
Out[1]:         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
```

The mean of the data regarding the Congruent list is 14.0511 and its standard deviation is 3.5594. The mean of the data regarding the Incongruent list is 22.0159 and its standard deviation is 4.7971.

(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 [26]: plt.hist(df['Congruent'], alpha=0.5, label='Congruent')
         plt.hist(df['Incongruent'], alpha=0.5, label='Incongruent')
         plt.legend(loc='upper right')
         plt.xlabel('Reading times')
         plt.ylabel('Frequency')
         plt.title('Reading time for congruent and incongruent words lists')
         plt.show();
```



For most people, it took between 13 to 15 seconds to read the congruent words list and between 21 to 23 seconds to read the incongruent words list. It usually took more time to read the incongruent words list than the congruent words list.

(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.

An statistical test applicable to distinguish the proposed hypotheses is a dependent t-test for paired samples. This choice depends on these assumptions: - The sample size is below 30. - The sample has an unknown population standard deviation. - The same subject takes the test twice.

```
In [36]: # difference of samples
         difference = df['Congruent'] - df['Incongruent']

         # calculate mean difference
         d_bar = np.mean(difference)

         #standard deviation of the differences
         sd = difference.std(ddof=1)

         # standard error of the mean difference
         se = sd/np.sqrt(len(difference))

         # t-statistic
         t = (d_bar)/(se)

         t

Out[36]: -8.020706944109957
```

For an alpha level of 0.05, and 24 degrees of freedom, the t critical values are -2.064 and 2.064. Since our t-statistic is pass these critical values, we can reject the null hypothesis in favor of the alternative hypothesis.

(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!

I think that having to hold 2 opposite ideas in the brain at the same time is responsible for the effects observed. The brain needs to activate two seperate areas, the one in charge of color perception and then the one associated with word encoding; finally, the brain descriminates which of the two is more important in that scenario. Alternatives tasks that would result in a similar effect could be trying to listen to an audiobook and have a conversation at the same time or texting and driving at the same time.

Resources

- Videos of Data Analyst Nanodegree term 1 by Udacity.com
- https://en.wikipedia.org/wiki/Stroop_effect
- http://www.statisticshowto.com/probability-and-statistics/hypothesis-testing/t-score-vs-z-score/
- https://stackoverflow.com/questions/2324438/how-to-calculate-the-statistics-t-test-with-numpy