Stroop Effect

Test a Perceptual Phenomenon

Data Analyst Nanodegree:
Statistics
Report Prepared By:
Marissa Schmucker

June 7th, 2017

This report analyzes the Stroop Effect data shown in the stroopdata.csv file. Using Google Sheets and Docs, I was able to analyze and present the data shown here. The purpose of this report is to pose and accept or reject a hypothesis that can help shed light on the Stroop Effect.

We start our report with our conditions, variables, test type, and hypothesis.

Conditions:
- Congruent Words: Words match the color
- Incongruent Words: Words do NOT match the color

Variables:
- Dependent Variable: Time it takes to name the ink color of each word
- Independent Variable: Conditions of each list (change in the color of each word)

Test Type:
- Dependent sample t-test
- Repeated Measures Design (amount of time recorded using each colored word list)
- A one-tailed positive t-test will determine whether the change in color (incongruent condition) will significantly increase the time taken to say the color of the letters in the word or whether there will be no significant change in time.

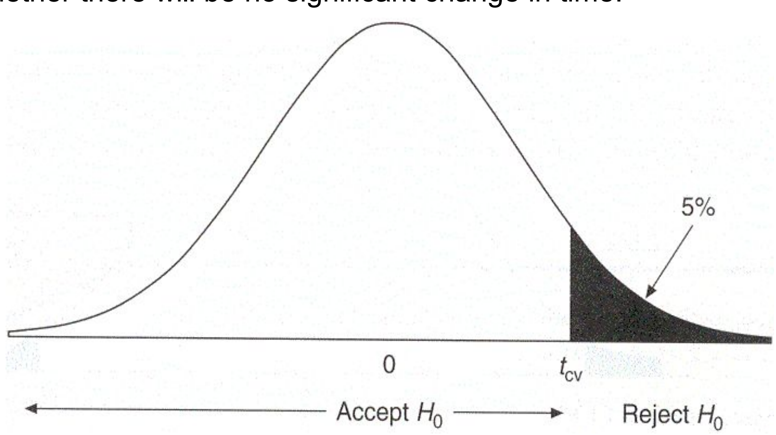


*image taken from Google Images
http://www.soes.soton.ac.uk/teaching/courses/oa132/module10/m10f53.gif

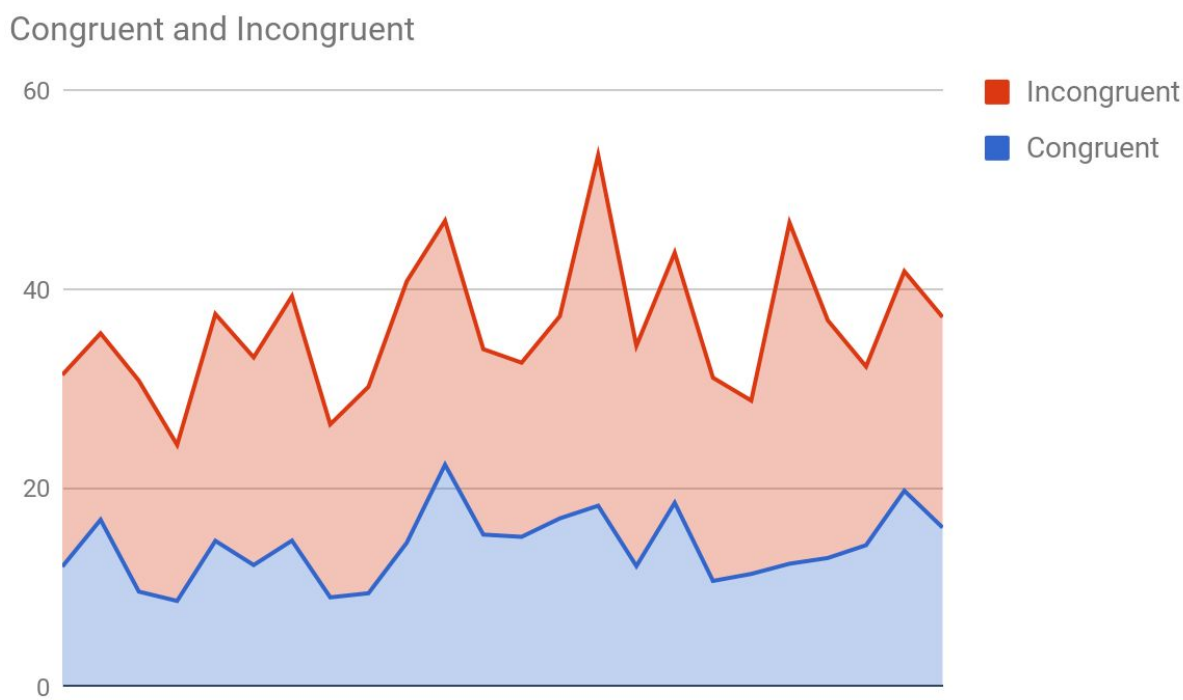
Hypothesis:

- Null Hypothesis(H_o): $\mu_c = \mu_i$

  There is no difference between the time it takes subjects to name each color in the first list (μ_c), containing congruent words, and the time it takes to name each color in the second list (μ_i), containing incongruent words.

- Alternative Hypothesis(H_a): $\mu_c < \mu_i$

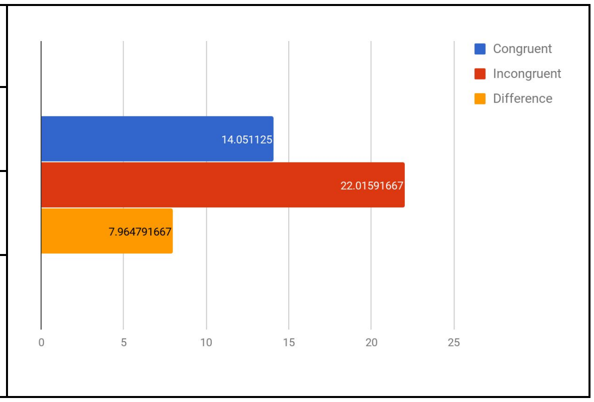
  There is a significant positive difference between the time it takes subjects to name each color in the first list and the time it takes to name each color in the second list.



As you can see from the chart above, there does seem to be a positive difference between time spent on the congruent words and the time spent on the incongruent words. We must continue to analyze the data if we hope to find if there is a significant difference. Let's start out with a few stats and then get to the good stuff, our conclusion. Will we accept the null hypothesis or reject it?

Moving right along…

Mean of Samples and Difference
μ_c = 14.051125
μ_i = 22.01591667
μ_d = 7.964791667



It looks like we have a relatively large difference in means. Let's see what our t-critical (t_c) value is before we find our standard deviation of differences (s) and solve for *t*. Since we are using a one-tailed positive t test, we will only reject the Null (H_o) Hypothesis if our *t* value is greater than our t-critical value with a percentage (p) greater than 0.05.

Looking at the t-table, we can see that with an alpha level of .05 and 23 degrees of freedom (n-1), our t-critical value will be 1.714.

$$t_c = 1.714$$

Now, let's find s and solve for t. We're almost there!

We find s by taking the difference between each sample difference and the difference of means, square the sum of all differences and divide by n-1 (df.)

$$s = \sqrt{\frac{(x_i - \mu_d)^2}{n-1}}$$ This gives us <u>s = 4.762</u>

Now we can get our t value easily by taking the difference of means over s divided by the square root of n.

$$t = \frac{\mu_d}{s/\sqrt{n}}$$ This gives us <u>t = 8.193</u>

Conclusion

So, remember that our t-critical value is 1.714. When we solved for t, we got 8.193 (rounded by three decimal places.) What does this tell us?

Should we reject the Null? **YES!**

We reject the hypothesis that there was no significant positive change because our t value is greater than our t-critical value. That means there is a significant positive change in time between the two samples.

Now that we can reject the null hypothesis, let's find our confidence interval (CI.)

To find our 95% confidence interval, we must first get the t-critical values. With an alpha level of .025 and 23 degrees of freedom, we get ± 2.069. Therefore, our confidence interval would be between $\mu_d - 2.069(s/\sqrt{n})$ and $\mu_d + 2.069(s/\sqrt{n})$.

With each value being rounded to three decimal places, we get...

95%CI = (5.953,9.976)

Now we can see where, most likely, our population mean lies! Somewhere between 6 and 10.

Now, let's go ahead and measure the distance between our means (μ_d) with standard deviations (s). Taking the difference of means over our standard deviation will give us a distance of 1.672 standard deviations. This distance is our Cohen's d value.

We can now say that our incongruent words sample took longer to finish than the congruent words sample. Now we can pose the question, why? Well, I suppose we should leave it up to the neuroscientists to answer that question. It's interesting to think about, though. I wonder what kind of results we would get if we did a similar test using numbers typed out with the number of dots associated with that number inside or around the number itself. Then have a second list of numbers with a random number of dots not associated with each given number. Each subject would have to record the amount of time it takes them to say the number of dots in each printed number. Hmmm… I wonder what that data would look like. I guess that's for another report.

Resources used to complete this report:

- Google image search
- http://www.wikihow.com/Write-a-Statistical-Report
- Google Sheets and Docs
- Udacity Statistics Course Videos and Resources, including the t-table (https://s3.amazonaws.com/udacity-hosted-downloads/t-table.jpg)