Nicholas Cica
Project 1
Data Analyst Nanodegree
August 2, 2016

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.

1. What is our independent variable?

The independent variable is the "word condition."

What is our dependent variable?

The dependent variable is the time is takes to name the ink colors

2. What is an appropriate set of hypotheses for this task?

Definitions:

H_O : *NULL Hypothesis*

H_A : *Alternative Hypothesis*

μ_I : mean of the time is takes to name the incongruent words

μ_C : mean of the time it takes to name the congruent words

Null Hypothesis:

The incongruent and congruent words mean are not different.

H_o: $\mu_I = \mu_C$

Alternative Hypothesis:

The incongruent words mean is greater than the congruent words mean.

H_A: $\mu_I > \mu_C$

What kind of statistical test do you expect to perform?

I will perform a two-tailed dependent t-test for paired samples because there are two conditions and this particular t-test tests whether the mean of the differences between dependent or paired observations are equal. I will not perform a z-test because I have less than

30 samples. Additionally, I do not know the population parameters such as population standard deviation.

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

Congruent

Central Tendency:

- Mode: N/A - No value occurs twice
- Median: 14.36
- Mean: 14.05

Measure of Variability

- Range:13.698
- Variance: 12.67
- Standard Deviation: 3.56

Incongruent

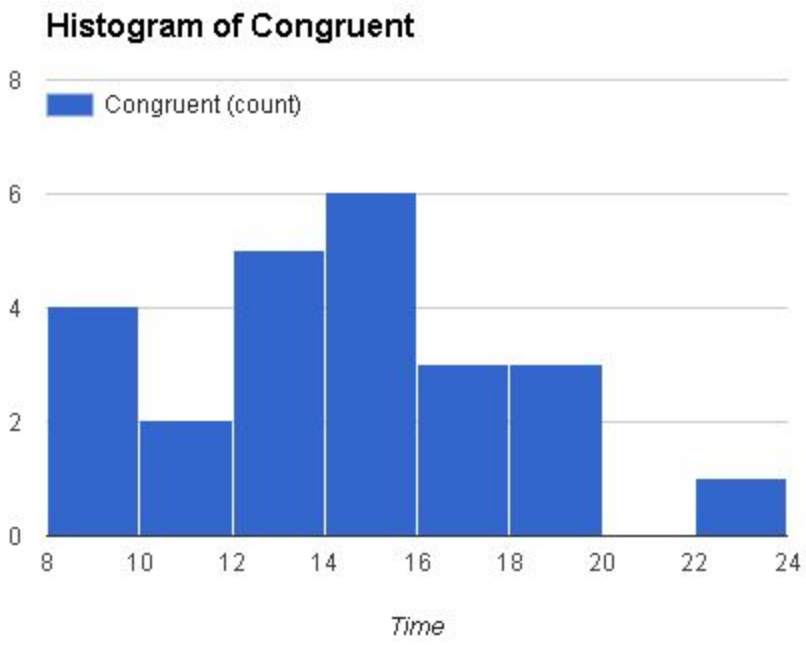
Central Tendency:

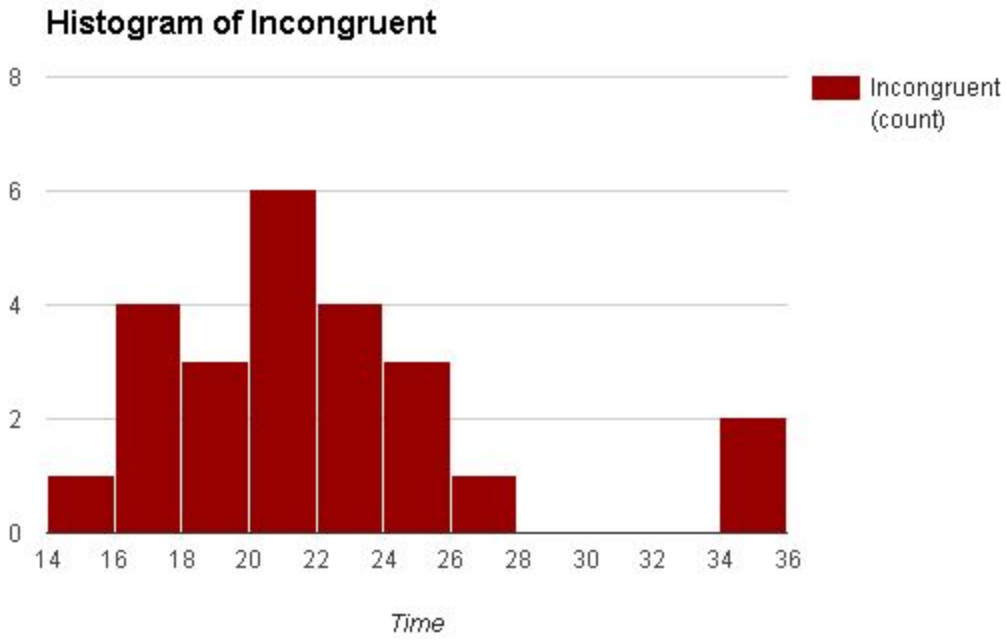
- Mode: N/A - No value occurs twice
- Median: 21.02
- Mean: 22.02

Measure of Variability

- Range:19.57
- Variance:23.01
- Standard Deviation: 4.8

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.





According to the visualizations, the average time to name the colors is higher for the incongruent words. Both histograms show one possible outlier.

5. Results

What is your confidence level and your critical statistic value?

Confidence interval on the mean difference; 95% CI = (-10.01, -5.92)

T-Statistic: -8.02
Alpha : 0.5
Df: 23
P-value: 0.0001
T-Critical Value: -2.064

t(23) = -8.02, p= 0.0001, two-tailed test

d = -1.64

Do you reject the null hypothesis or fail to reject it?
Reject the NULL hypothesis since the T-Statistic is past the critical value in the critical region.

Come to a conclusion in terms of the experiment task.
We can conclude that the incongruent words require more time to name than the congruent words. Participants will spend on average between 4 and 11 second fewer on the congruent words.

Did the results match up with your expectations?
Yes, the results match my own expectations based on my own experiences with the experiment.

References:
http://stattrek.com/descriptive-statistics/variability.aspx?Tutorial=AP
http://www.pryor.com/blog/3-easy-formulas-to-calculate-the-range-of-values-in-excel/
http://support.minitab.com/en-us/minitab/17/topic-library/basic-statistics-and-graphs/hypothesis-tests/tests-of-means/types-of-t-tests/