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?

Independent variable: the independent variable is the specific condition in which the words are displayed, *i.e.* congruent words condition and incongruent words condition. This is the variable upon which the researcher has control.

Dependent variable: the dependent variable is the time it takes to name the ink colors in which the words are written. The time needed for the task of naming the ink colors is presumably dependent on the task condition (the independent variable).

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

Hypotheses: the time required to name the ink colors depends on the task's condition. Presumably the time needed to name the ink colors is higher in the incongruent words condition (this has been actually confirmed experimentally by John Ridley Stroop in 1935 [1] and many other subsequent studies). This is because the brain has to resolve a conflict between the semantic of the word and the visual perception of the ink color in which the word is written.

To interpret the result of the experiment I would resort to the Wilcoxon signed-rank test, a non-parametric statistical method used to test for consistent differences between pairs of observations [2]. Given two samples X and Y of n paired observations (x_i, y_i : i = 1, …, n), the null hypothesis is that the median of X and the median of Y are statistically the same.

Because of what discussed before we expect that the reaction-time in the incongruent condition is higher (if different) than the reaction-time in the congruent condition. For this reason the most appropriate form of the Wilcoxon signed-rank test is the *one-tailed* version, with a critical p-value p* of 0.05 [3-4].

The reason for using a non-parametric test is that the shape of the distributions to be

compared (*i.e.* the distributions of the reaction-time in the two conditions) is not known *a priori*. Also, the distributions (although of unknown shape) satisfy the requirements of the Wilcoxon signed-rank test. In particular, if we denote the difference y_i - x_i with z_i, then:

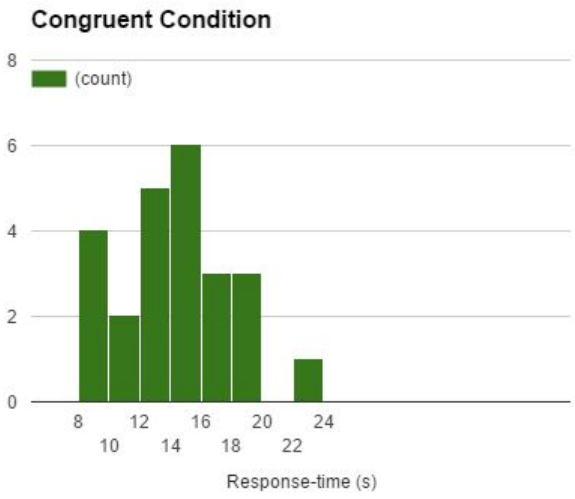
1. z_i are independent across the value of i = 1, .., n;
2. each z_i comes from the same continuous population;
3. the values x_i and y_i are ordered (*i.e.* the comparisons "greater than", "less than", and "equal to" are meaningful).

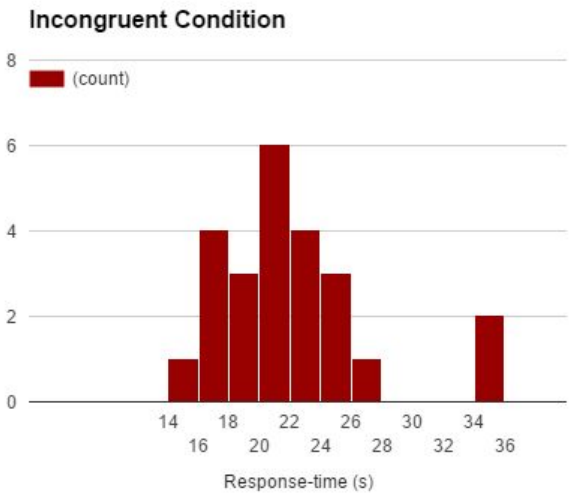
Now, download this dataset which contains results from a number of participants in the task.

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

	Congruent	**Incongruent**
Mean	14.05	22.02
St. dev.	3.56	4.80
Median	14.36	21.02
1st quartile	11.90	18.72
3rd quartile	16.20	24.05
min	8.63	15.69
max	22.33	35.26

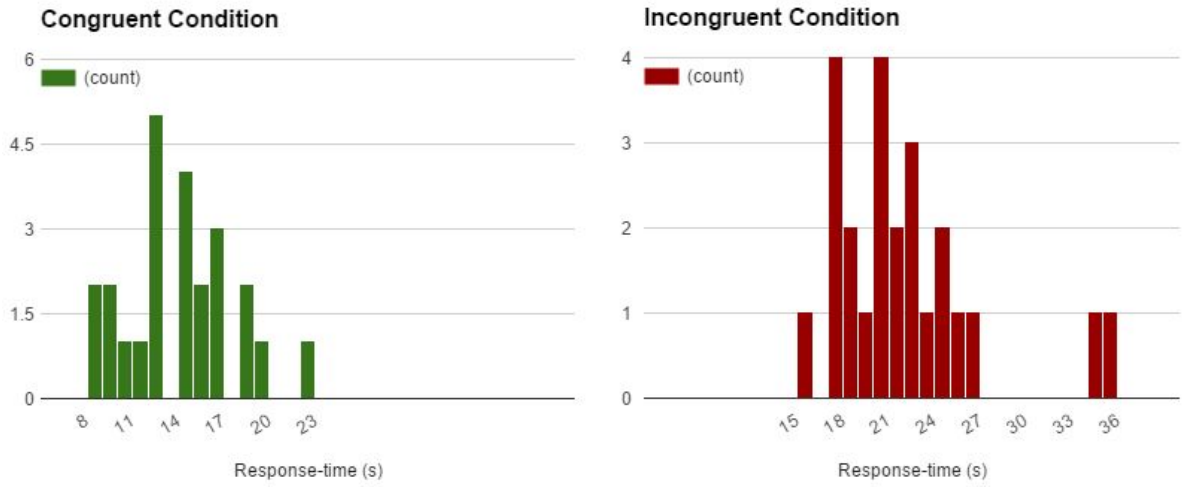
4. Provide one or two visualizations that show the distribution of the sample data. Write one or two sentences noting what you observe about the plot or plots.





The histograms of the response-time for the incongruent condition is characterized by larger values of the variable compared to the distribution obtained for the congruent condition.

Interestingly both distributions present a gap between the main bulk of the sample and those data-points with the largest value of the response-time. It is not clear whether this is due to an artifact in the data-set that would allow us to class the isolated data-points as outliers. A possibility is that there was a sub-set of people participating in the experiment who had a particularly slow response-time, and this was reflected in both conditions. On the other hand, we also notice that when decreasing the binwidth of the histograms, more gaps appear in both distributions. This shows that the original gap was at least partly an artifact due to the particular value of the binwidth used before (this observation mostly applies to the congruent condition though).



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?

The p-value obtained from the one-tailed Wilcoxon signed-rank test is 9.12×10^{-6}, several orders of magnitude lower than the critical value $p^* = 0.05$. The null hypothesis was therefore rejected, *i.e.* the response-time of the incongruent condition is statistically higher than the response-time of the congruent condition with more than 95% confidence.

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!

The reason for the incongruent condition to be associated with longer response-times may lay in the fact that processing words happens in a more automatic fashion in our brain than the more mentally effortful task of naming the ink color. Different theories have been proposed to explain why this would be the case: *processing speed*, *selective attention*, *automaticity* and *parallel distributed processing*. All these theories are commonly known as ‘race models’. [5]

The task of making an appropriate response when given two conflicting signals has tentatively been located in a part of the brain called the *anterior cingulate*. Broadly

speaking the *anterior cingulate* acts as a conduit between lower (somewhat more impulse-driven) brain regions and higher (somewhat more thought-driven) behaviors. The Stroop effect's sensitivity to changes in brain function may be related to its association with the *anterior cingulate* [6].

A similar task that would result in a similar effect is a variation of the original Stroop test called *numerical Stroop test*. The numerical Stroop test demonstrates the close relationship between numerical values and physical sizes. In this test a digit can be presented as big or small (e.g., 5 vs. 5), irrespective of its numerical value. Comparing digits in incongruent trials (e.g., 3 5) is slower than comparing digits in congruent trials (e.g., 3 5). The difference in reaction time is termed the *numerical Stroop effect*.

The response time of comparing two numbers by physical size is affected by the numerical distance between the members of the digit pair, indicating that numerical distance is automatically computed even when it is irrelevant to the comparative judgment being required by the task [7].

Bibliography

[1] Stroop, J. Ridley (1935), *Studies Of Interference in Serial Verbal Reactions*. Journal of Experimental Psychology, *18*, 643-662.

[2] *Wikipedia*. Wilcoxon signed-rank test.

[3] *Wikipedia*. One- and two-tailed tests.

[4] GraphPad Statistics Guide.

[5] *Wikipedia*. Stroop effect.

[6] *Oracle* | *Maximizer*. The Difference Between One-Tailed & Two-Tailed Testing.

[7] Avishai Henik, Joseph Tzelgov (1982), Is three greater than five: The relation between physical and semantic size in comparison tasks. Mem Cogn Memory & Cognition - 07/1982.