Project 1:
Statistics: The Science of Decisions

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

Our independent variable is the list of words that are presented in both conditions (congruent and incongruent) to the readers.
Our dependent variable is the time that it takes the reader to name the ink color of the list for each list in each condition.

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

Hypothesis:
Null hypothesis: (H_0) The mean of time that it takes for the readers to say out loud the ink color of the list of congruent words ($\mu_{congruent}$) is the same as the mean of time that it takes them to say the ink color of the list of all incongruent words ($\mu_{incongruent}$).
Mathematical expression for H_0: $\mu_{congruent} = \mu_{incongruent}$

Alternative hypothesis: (H_1): The mean of time that it takes for the readers to say out loud the ink color of the list of congruent words ($\mu_{congruent}$) is totally different from the mean of time that it takes them to say the ink color of the list of all incongruent words ($\mu_{incongruent}$).
Mathematical expression for H_1: $\mu_{congruent} \neq \mu_{incongruent}$

The statistical test that I should perform here is a two-tailed dependent t test. This is because the same subjects are tested under two different conditions, therefore I think the choice of a 'dependent' test is an appropriate choice here. A t test is preferred over z test is because we don't really know the population parameters. Finally, I chose to conduct a two-tailed t test instead of a one-tailed t test because I don't really care about the direction of the alternative hypothesis, meaning it does not matter to me which list will take more time for the readers to say out loud all the words.

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

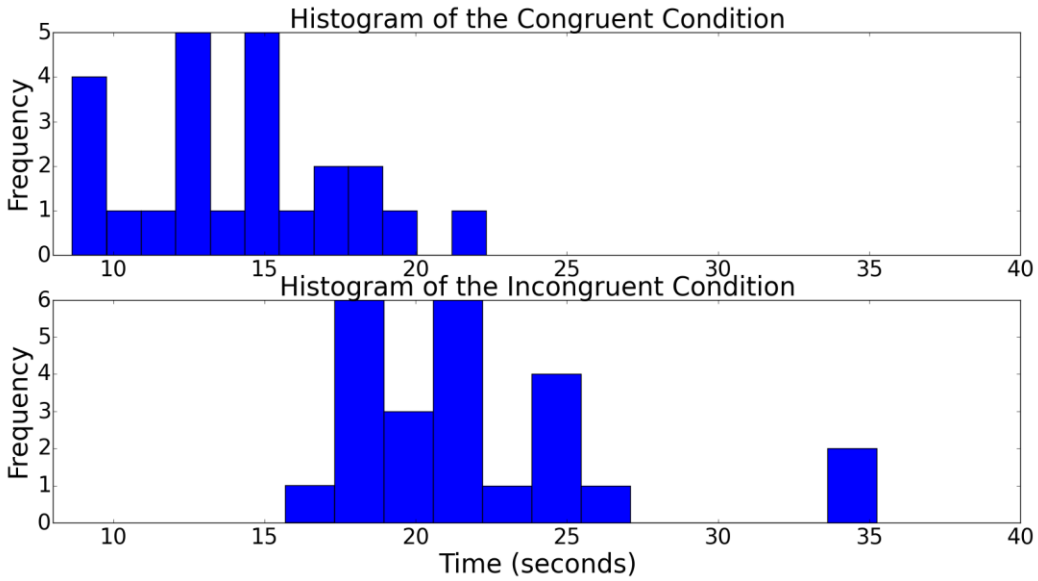
	Mean	***Median***	***Mode***	***Variance***	***Standard Deviation***
Congruent	14.05	14.36	#N/A	12.67	3.56
Incongruent	22.02	21.02	#N/A	23.01	4.80

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.

Below are the histograms for the congruent condition (top) and the incongruent condition (bottom). The x axis is in unit of time (seconds) and the y-axis represents the frequency. According the two histograms, on the same x axis scale, it generally takes more time for the readers to completely say

Project 1:
Statistics: The Science of Decisions

out loud all the ink color of the list of all the incongruent words than for the list of the congruent words.



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?

Null hypothesis: H_0: $\mu_{congruent} = \mu_{incongruent}$
Alternative hypothesis: H_1: $\mu_{congruent} \neq \mu_{incongruent}$

$$\mu_{congruent} = 14.05$$
$$n_{congruent} = n_{incongruent} = 24$$
$$\mu_{incongruent} = 22.02$$
$$s_{congruent} = 3.56$$
$$s_{incongruent} = 4.80$$
$$\mu_D = \mu_{congruent} - \mu_{incongruent} = -7.97$$

The difference in standard deviation,

$$s_D = 4.86$$

The t-statistic is:

$$t_{stat} = \frac{\mu_D}{\frac{s_D}{\sqrt{n}}} = -\frac{7.97}{\frac{4.86}{\sqrt{24}}} = -8.03$$

With α = 0.05, degree of freedom = 23, the $t_{critical} = \pm 2.069$

Since the $t_{stat} < t_{critical}$, I would reject the null hypothesis. Therefore, there is a statistical significant difference between the time it took to say out loud the words in the congruent condition

Project 1:
Statistics: The Science of Decisions

and the time to say all words in the incongruent condition. This result does match up with my original expectation.
With 95% confidence, the confident interval (CI) is:

$$\mu_D \pm t_{critical} * \frac{s_D}{\sqrt{n}} = (-7.97 - 2.069 * 0.99, -7.97 + 2.069 * 0.99)$$

$$CI = (-10.02, -5.92), \alpha = 0.05$$

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!

One of the main reasons why it took more time for say out loud the incongruent words is because the brain cannot help but read. Therefore, most readers will tend to say the word that they see rather than the color of that word. When there is a conflict between the two sources of information, the word itself and the color of ink that it is printed, our cognitive load is increases. Thus, our brain has to work harder to solve the problem. Performing this extra step takes more time and slow down our responses.

Another task that is in similar effect would be the number Stroop effect, where two numbers are presented. However, the smaller number will be printed in a bigger font-size than the larger number. The readers will be asked to say out loud which number is the bigger number based on what they see.