Christian Rivera
Udacity Project 1
Statistics: The Science of Decisions

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

a. The Independent variables are the lists of word/color combinations.
b. The dependent variable is the amount of time each participant takes to read through the list.

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

Null hypothesis: there is no statistical difference between the time it takes for a person to complete the congruent test and the time it takes him for the incongruent test. This is shown in that the mean average time to complete the congruent test is equal to the mean average time for completing the incongruent test.

$H_0: \mu_1 = \mu_2$

Alternative hypothesis: Due to the increased complexity caused by the disconnect of the color of the word not matching the content of the word, the incongruent test takes more time to complete than the congruent test. This is shown as the mean average time of completion for the incongruent test is statistically longer than the mean average completion time for the congruent test.

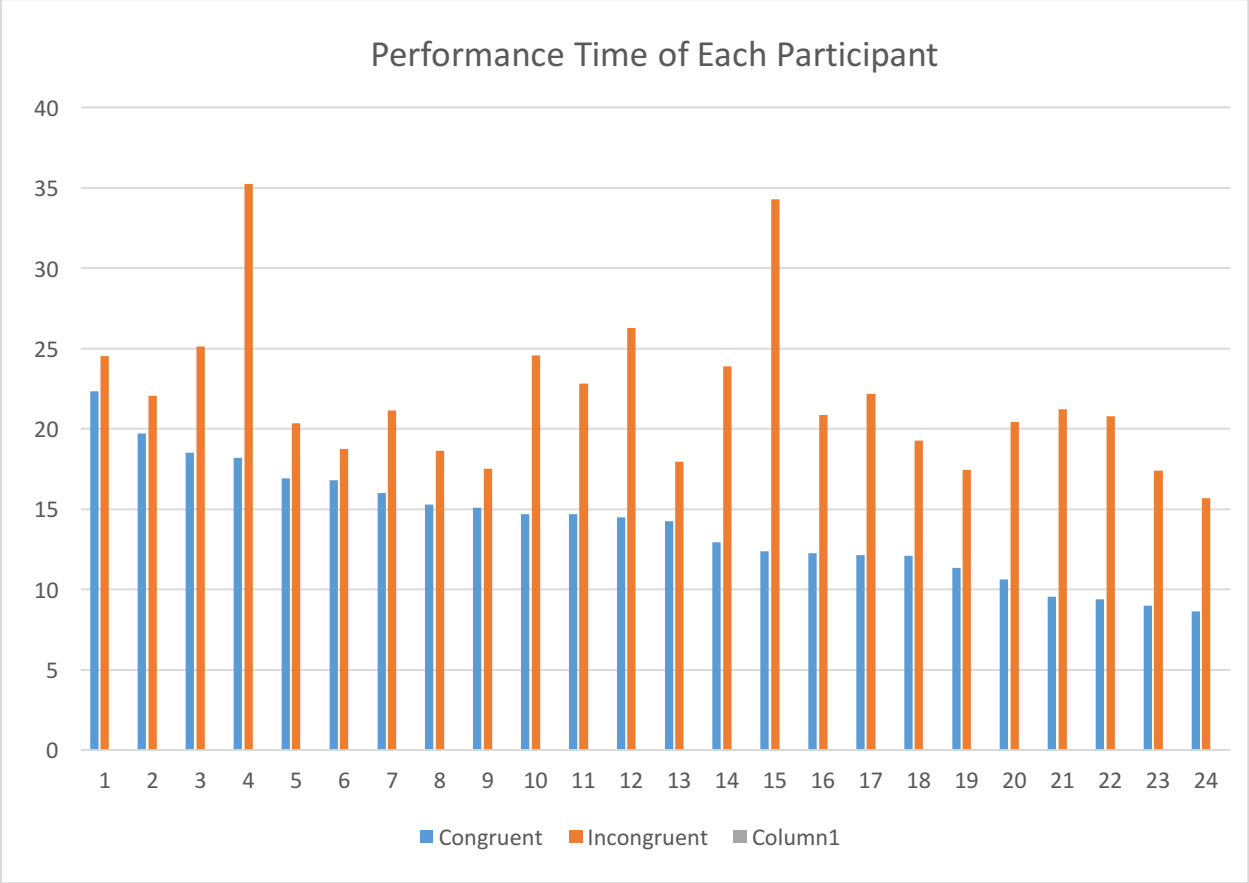
$H_1: \mu_1 < \mu_2$

Statistical Test: To determine if there is a difference between the time it takes analyzing the congruent and non-congruent word/color combinations, I will use the dependent t-Test as the same subject group is used for both congruent and incongruent tests, the sample size is below 30 (24), and because I do not know the standard deviation of the population

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
Median	14.36	21.02
Variance	12.67	23.01
Std Deviation	3.56	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.*



From this graph, we can see that every participant took more time on the incongruent test than on the congruent test. We can also tell that there is no correlation between completion time on the congruent test and incongruent test as the fastest congruent test takers did no better on the incongruent test than the slower congruent test participants.

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?*

1-tailed Alpha = 0.05
Degrees of freedom = N – 1 = 24 – 1 = 23
Critical Value = 1.714

If the absolute value of t is greater than 1.714, then I reject the null hypothesis.

Step 1: Compute the differences between the two data sets:

Congruent	Incongruent	Difference
22.328	24.524	-2.196
19.71	22.058	-2.348
18.495	25.139	-6.644
18.2	35.255	-17.055
16.929	20.33	-3.401
16.791	18.741	-1.95
16.004	21.157	-5.153
15.298	18.644	-3.346
15.073	17.51	-2.437
14.692	24.572	-9.88
14.669	22.803	-8.134
14.48	26.282	-11.802
14.233	17.96	-3.727
12.944	23.894	-10.95
12.369	34.288	-21.919
12.238	20.878	-8.64
12.13	22.158	-10.028
12.079	19.278	-7.199
11.344	17.425	-6.081
10.639	20.429	-9.79
9.564	21.214	-11.65
9.401	20.762	-11.361
8.987	17.394	-8.407
8.63	15.687	-7.057

The mean of the difference column = -7.96.

Step 2: determine the standard deviation:

$$s_D = \sqrt{\frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1}}$$

Variance = (2066.84 – 36540.23/24)/23 = 23.67

Standard deviation = sqrt(23.67) = 4.86

Step 3: determine t-score:

t-score = -7.96 / (4.86 / sqrt(24)) = -8.02

t-score = -8.02

Since the t-score is -8.02 and -8.02 is less than the critical value, -1.714, I can reject the null hypothesis.

From the data, it is clear that there is a statistical difference between the time it takes to complete the congruent test and the time it takes to complete the incongruent test. This falls in line with my personal experience taking the two tests as well as the reasoning behind the alternative hypothesis.