Statistics: The Science of Decisions

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.

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

Our dependent variable is the amount of time that it takes to answer
Our independent variable are congruent and incongruent conditions

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

NULL hypothesis (H_0): the average time it takes to respond congruent and incongruent responses is equal.

H_0: $\mu C = \mu I$ (the mean of the congruent responses is equal to the mean of incongruent responses)

Alternate hypothesis (H_1): the average time it takes to respond congruent and incongruent responses is different.

H_1: $\mu C \neq \mu I$ (the mean of the congruent responses is not equal to the mean of incongruent responses)

What kind of statistical test do you expect to perform? T-test. I do this because we don't have the standard deviation of the population but also because our sample size (n) is less than 30. The type of t-test I used is a 2 Sample t-test called paired t-test, to find out whether the mean of the differences between of two paired samples is zero or not equal to zero. Also, to calculate a range of values that is likely to include the population mean of the differences

We are going to run the analysis with the dataset provided by Udacity. It contains results from a number of participants in the task (n=24). Each row of the dataset contains the performance for

one participant, with the first number their results on the congruent task and the second number their performance on the incongruent task.

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

First, we analyze the Congruent data:

summary(dfstroop$Congruent)

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
8.63	11.90	14.36	14.05	16.20	22.33

Standard Deviation is 3.559358
sd(dfstroop$Congruent)

Variance is 12.66903
var(dfstroop$Congruent)

Now, we analyze the Incongruent data:

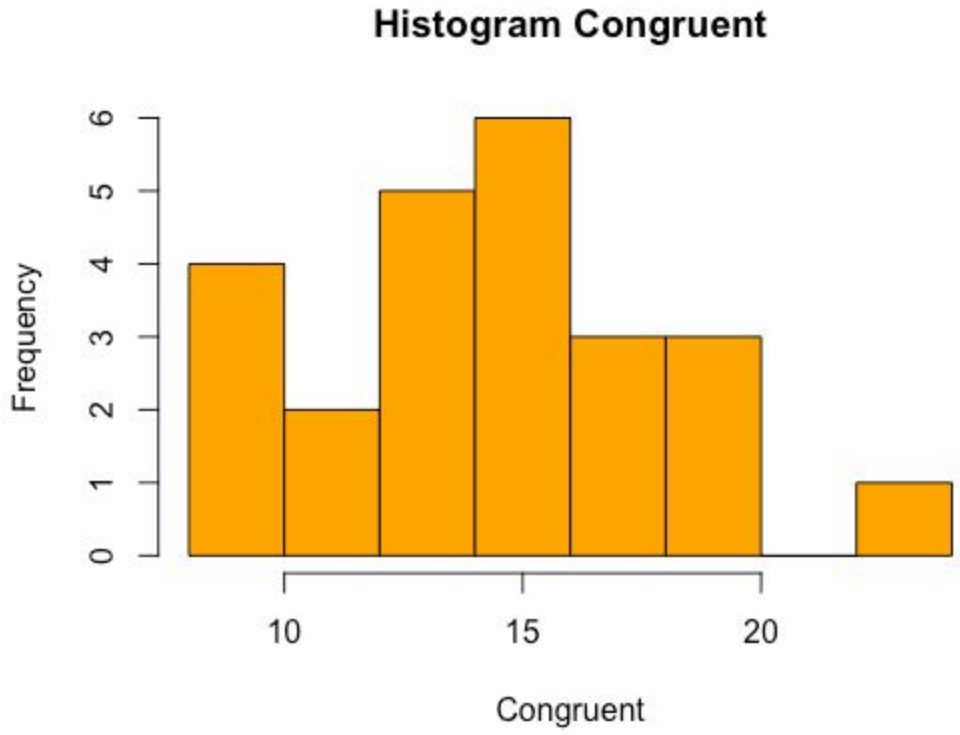
summary(dfstroop$Incongruent)

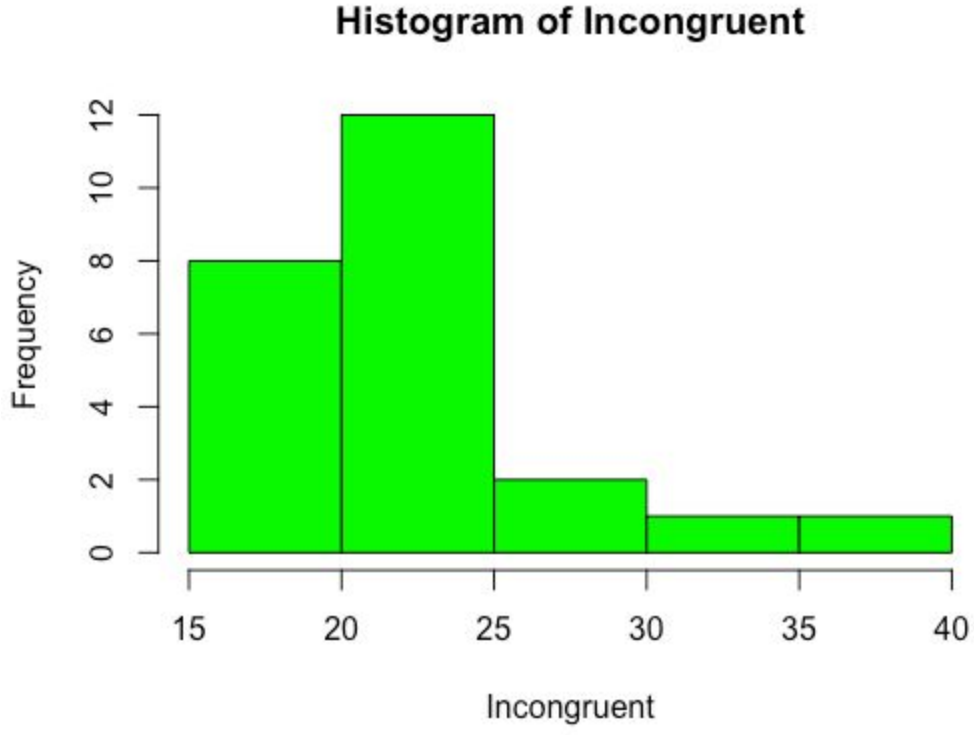
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
15.69	18.72	21.02	22.02	24.05	35.26

Standard Deviation is 4.797057
sd(dfstroop$Incongruent)

Variance is 23.01176
var(dfstroop$Incongruent)

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.

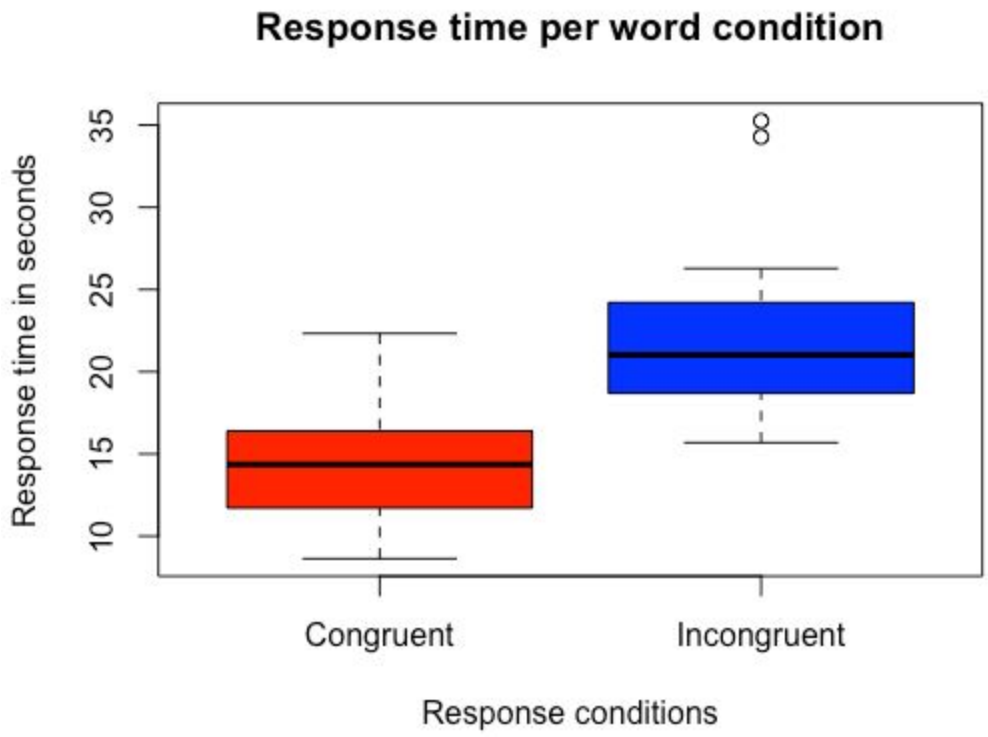

Histogram Congruent
Frequency
0
1
2
3
4
5
6
10
15
20
Congruent


Histogram of Incongruent
Frequency
0
2
4
6
8
10
12
15
20
25
30
35
40
Incongruent

What we observe here is that the average time it takes to respond to a congruent question is higher than the average time it takes to respond an incongruent question. We can also see that in the median values for each.

Also, most of the congruent responses took between 12-16 seconds to respond, while most of the incongruent responses took between 20-25 seconds.

I will use boxplot to further analyze the samples.



We can clearly see how the means (and the medians) differ, but also we observe a couple of outliers in the Incongruent group. Here is the code I wrote:

boxplot(database, col = c("red", "blue"), main = "Response time per word condition", ylab = "Response time in seconds", xlab = "Response conditions")

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?

I want to do a T-Test to determine whether the means of two groups are equal to each other.

I used the sample dataset provided in this exercise. Then, added a new variable (column) to my dataset called (difference). It shows the difference between the Congruent and the Incongruent responses. Here is the code I used:

```
dfstroop$difference <- dfstroop$Congruent - dfstroop$Incongruent
```

Then, I run a summary for that new column, so that I can obtain the mean of the difference.

summary(dfstroop$difference)

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-21.919	-10.258	-7.667	-7.965	-3.646	-1.950

I also calculated the standard deviation

sd(dfstroop$difference)
Standard deviation of the difference = 4.864827

Now, I would be able to calculate the T-Test by using the formula, but I am going to take a shortcut. Using R here:

```
t.test(dfstroop$difference)
```

Here is the result:

```
t = -8.0207, df = 23, p-value = 4.103e-08
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
-10.019028  -5.910555
sample estimates:
mean of x
-7.964792
```

The t Score is: **2.069.** Since my t statistic is -8.0207, which is significantly less than my t-score, I reject the Null Hypothesis that the time it takes to respond congruent and incongruent responses is equal. I did expect this result.

6. Optional: What do you think is responsible for the effects observed?

The test requires the brain to inhibit part of the information presented to us, in this case the word meaning, in favor of the color of the word. When both match, our brain easily responds, but when they don't match, our brain needs to adapt and inhibit part of the stimulus.

Can you think of an alternative or similar task that would result in a similar effect?

This happens when I am talking and at the same time someone is talking to me. I need to inhibit the voice of the person talking to me to be able to talk. Which is challenging at times...