STROOP EFFECT-The test of a perceptual phenomenon.

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

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

Congruency (Congruent/Incongruent) of the words in color as reported by the individuals is an Independant Variable

The Time taken to detect the words in a given list is the dependant 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.

$H_0 : \mu c - \mu i = 0 ;$

$H_A : \mu c - \mu i \; != 0$

Where, H_0 represents the Null Hypothesis which says that there will not be any significant change. In other words the mean difference of time in both conditions will be same.
H_A is the Alternative Hypothesis which says that there is a significant difference in the means of time at two conditions (Congruent/Incongruent words) and it is either larger or smaller. μc is the Mean of the values of times recorded in reading the Congruent words. μi is the Mean of the times in Incongruent list of words.

I would like to validate a **'Two-Tailed dependant sample t Test at 95% Confidance Interval'** as I chose to test the alternative hypothesis because, I have a strong feeling that atleast in one of the cases the time taken for reading the words in both conditions will be different. Using the t values as the sample we have here is limited(less than 30) and we do not know the Mean or the Variance of the population.

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

Stroop Data, Times recorded in Seconds

Congruent	Incongruent
12.079	19.278
16.791	18.741
9.564	21.214
8.630	15.687
14.669	22.803
12.238	20.878
14.692	24.572
8.987	17.394
9.401	20.762
14.480	26.282
22.328	24.524
15.298	18.644
15.073	17.510
16.929	20.330
18.200	35.255
12.130	22.158
18.495	25.139
10.639	20.429
11.344	17.425
12.369	34.288
12.944	23.894
14.233	17.960
19.710	22.058
16.004	21.157

```
[1] "The number of rows representing the number of samples, n = 24"

   Congruent      Incongruent
 Min.   : 8.63   Min.   :15.69
 1st Qu.:11.90   1st Qu.:18.72
 Median :14.36   Median :21.02
 Mean   :14.05   Mean   :22.02
 3rd Qu.:16.20   3rd Qu.:24.05
 Max.   :22.33   Max.   :35.26
```

Sample size, n = 24

Mean for Congruent words= 14.05 sec and for Incongruent words = 22.02 sec
Median (which is a better reperesentation of central tendancy) for Congruent words = 14.36 sec and for Incongruent words = 21.02 sec

Now, let us calculate the Standard Deviation in each case.

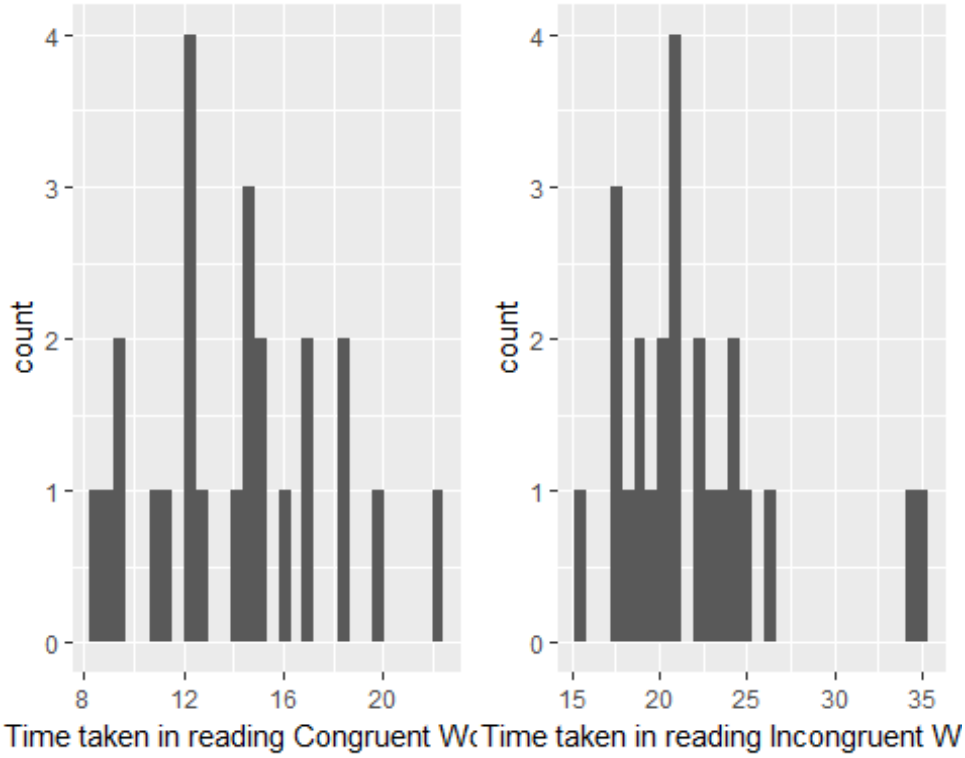
The Standard Deviation, SD is calculated using the formula:

SD = Sqrt $(\sum(x_i - \mu)^2 /n)$

where,
x_i = Each value of time recorded in sec,
μ is the Mean value of the times,
n = number of values in the list.

```
[1] "The Standard Deviation of Congruent Values is 3.56 sec and Incongruent
values is 4.80 sec"
```

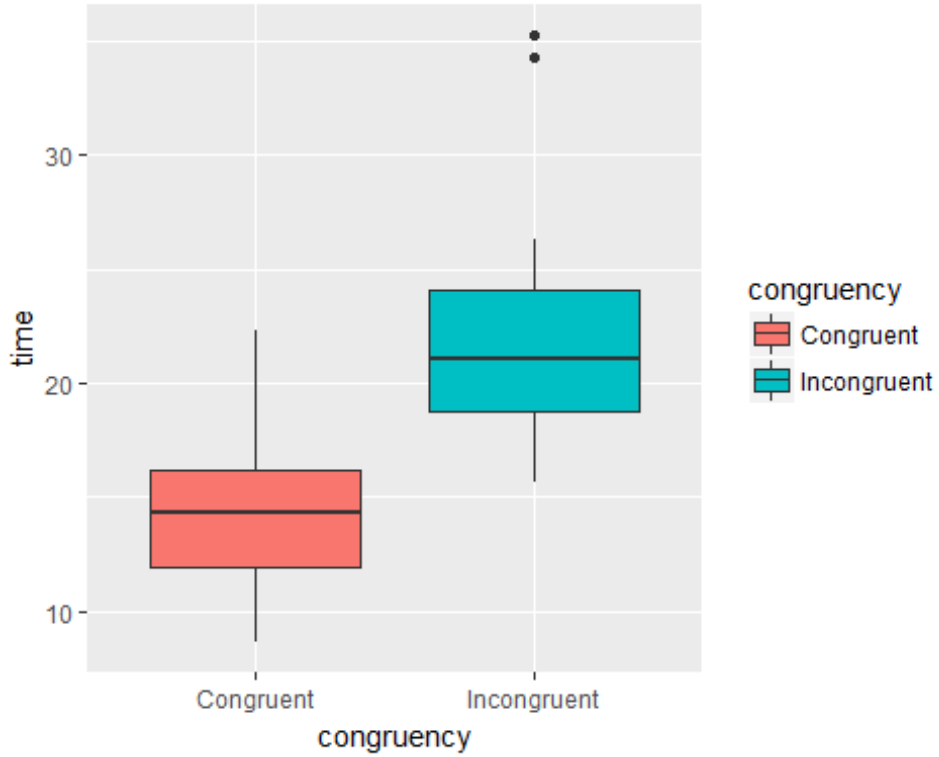
4. Provide one or two visualizations that show the distribution of the sample data. Write one or two sentences noting what you observe about the plot or plots.



These simple histograms show us that the distribution of the time taken to read the congruent words are slightly right skewed. However, the distribution of the time taken to read the Incongruent words is more normalised with clear outliers.

```
# A tibble: 2 x 5
  congruency  `mean(time)` `median(time)` `sd(time)` `var(time)`
```

	<chr>	<dbl>	<dbl>	<dbl>	<dbl>
1	Congruent	14.1	14.4	3.56	12.7
2	Incongruent	22.0	21.0	4.80	23.0



This box plot shows the significant difference in the medians of time taken in reading under the two conditions (Congruent and Incongruent). The time taken to read the incongruent words by each person is more than the time taken to read 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?

The difference of Congruent and Incongruent times sample

x
7.199
1.950
11.650
7.057
8.134
8.640
9.880
8.407
11.361

11.802
2.196
3.346
2.437
3.401
17.055
10.028
6.644
9.790
6.081
21.919
10.950
3.727
2.348
5.153

Now with the above sample, we will calculate the values of parameters to test our Hypothesis:

Hypothesis

$H_0 : \mu c - \mu i = 0$;
$H_A : \mu c - \mu i != 0$

The following formulae are used to calculate:

```
1) Calculate the difference between means.
     mean_diff = $\mu~c~$ - $\mu~i~$
2) Calculate the standard error of Means
     EM = $\s~d~$/$\sqrt(n)$, where $\s~d~$ = standard deviation of the
differences.
3) Calculate the *t-statistical value* :
     t = mean_diff / SEM

[1] "Sample size, N = 24\n Degree of Freedom, DF = 23\n Standard Error of
Means, SEM = 0.99\n t Statistical value, t = 8.02\n Mean Difference,
mean_diff = 7.96\n Difference in Standard Deviation, sd_diff = 4.86\n"

4) By using the *t-table* for degree of freedom, **DF (N - 1) = 23**, at
**95%** Confidence Interval and ** $\aplha$ = 0.025** in a two-tailed test,
we get a t Critical values, $^t^(*)$ = + 2.069 and - 2.069
```

Decision

Clearly, when t- statistical value is lying outside the t-critical values, beyond the α region, we have proved the Alternative Hypothesis, accepting the fact that there is a significant

difference between the time taken to read the congruent and Incongruent words. In other words we reject the null, which means $\mu c - \mu i$!= 0

So, Yes. The results were as expected in the beginning of the analysis.

We can further find the Confidence Interval for 95% as,
CI = (mean_diff - (t x SEM) , mean_diff + (t x SEM))

```
[1] "The lower bound for 95 percent CI is: 0.00\n The upper bound for 95
percent CI is: 15.93"
```

95% of Confidence Interval is (0, 15.93). This is by how much each time of reading can vary between reading the congruent and Incongruent words.

Cohen's d:

```
[1] "Cohen's d = 2.24"
```

For the case of the paired test, Cohen's d is defined as the difference of means between the post-test and pre-test treatments divided by the standard deviation of the pre-test condition. And it is calculated as:

Cohen's d = $\mu i - \mu c$ / SD_c = 2.24

Correlation Measures.

```
[1] "The proportion of variablity is 0.74"
```

The percentage of this variability is defined by r^2 and measured by the formula,
$r^2 = (t^2) / (t^2 + DF)$

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!

When the name of a color (e.g., "blue", "green", or "red") is printed in a color that is not denoted by the name (e.g., the word "red" printed in blue ink instead of red ink), the learned brain is confused naming the color of the word and takes longer than when the color of the ink matches the name of the color.

Alternative/supplimental tests of ths phenomenon:

```
Shape incongruency: Print words inside of the shapes they describe (circle,
square, triange, etc) in the congruent test, then have the words and shapes
be mismatched for the Incongruent test. This is a good alternative test that
when paired with the original Stroop test could help in the understanding of
whether or not word comprehension is the dominant mechanism here.

Audible incongruency: Test audio/visual incongruency by displaying an image
of an animal while having a recorded naming of that animal, and have the
subject name/make the sound made by the animal (testing not just repetition
of wha is told, but the audio/visual association with animal traits). Repeat
with incongruently displayed/spoken animals.
```

References

https://www.rstudio.com/wp-content/uploads/2016/03/rmarkdown-cheatsheet-2.0.pdf
http://www.statisticshowto.com/probability-and-statistics/hypothesis-testing/t-score-vs-z-score/ https://s3.amazonaws.com/udacity-hosted-downloads/t-table.jpg
https://en.wikipedia.org/wiki/Stroop_effect