

P1.

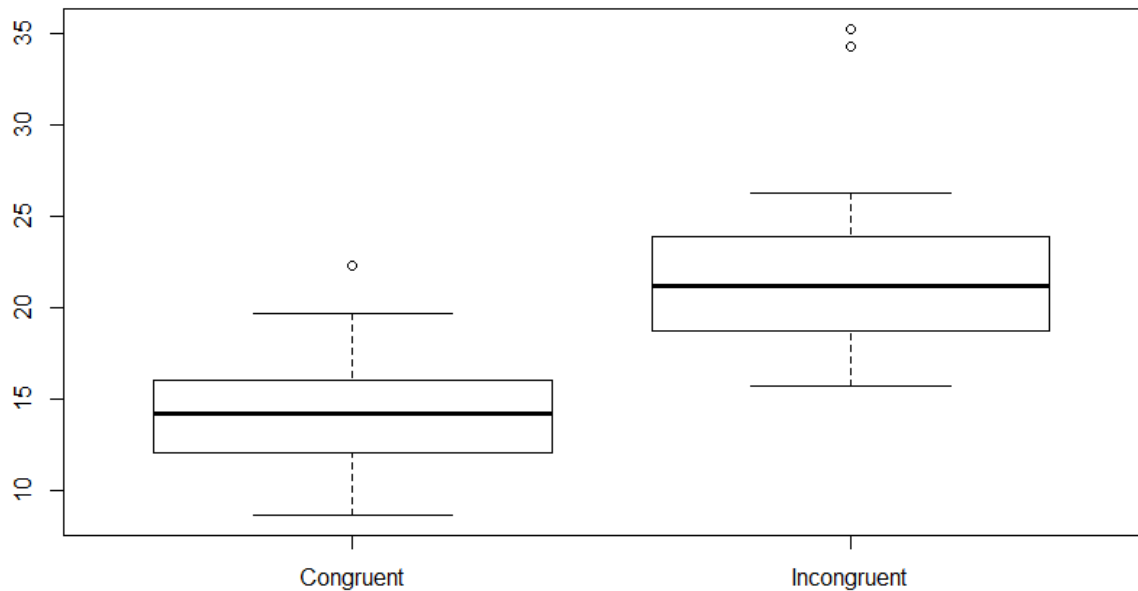
In this project, unless specified, I will use the statistical package-R to demonstrate all my works. I will also copy all the graphs from the output window of R.

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

Independent variable would be the two tests, or the actual words that were used in the both the congruent and incongruent tests.

The dependent variables are the time elapsed (in Seconds) in reading the words.

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



By simply using the box plot of the data, we can safely assume that, the mean time of incongruent is more than that of congruent words reading (The box plot depicted above, clearly shows our preference for the alternative hypothesis to be one tailed)

$H_0: \mu_i - \mu_c = 0$, there is no difference in the mean reading time of the two set of words (entire population)

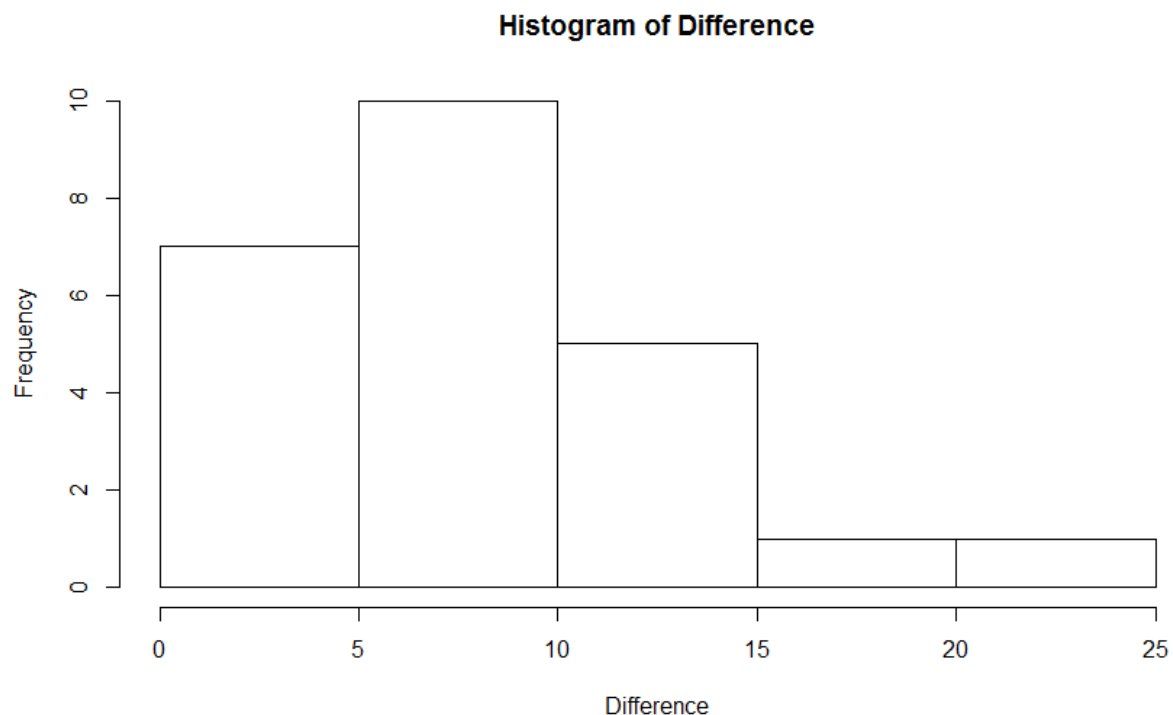
$H_1: \mu_i - \mu_c > 0$, the mean reading time of incongruent words is greater than that of congruent words

NB. μ_i = Mean of the incongruent population, μ_c = Mean of the Congruent population

We use the Paired-samples t-test (dependent t test) since we are measuring the reading time of the same person (that is each row in the two columns represent the reading time of one person).

Assumptions in conducting our test:

1. The data are continuous (Time is continuous data)
2. The difference in reading time for the matched-pairs, follow a normal probability distribution (the histogram for the difference supports our assumption)
3. Each pair of words have an equal probability of selection from the whole population.



The above graph shows the histogram for the difference of the individual reading time for the incongruent and congruent set of words. As we can see from the graph, it tends to be normal curve,

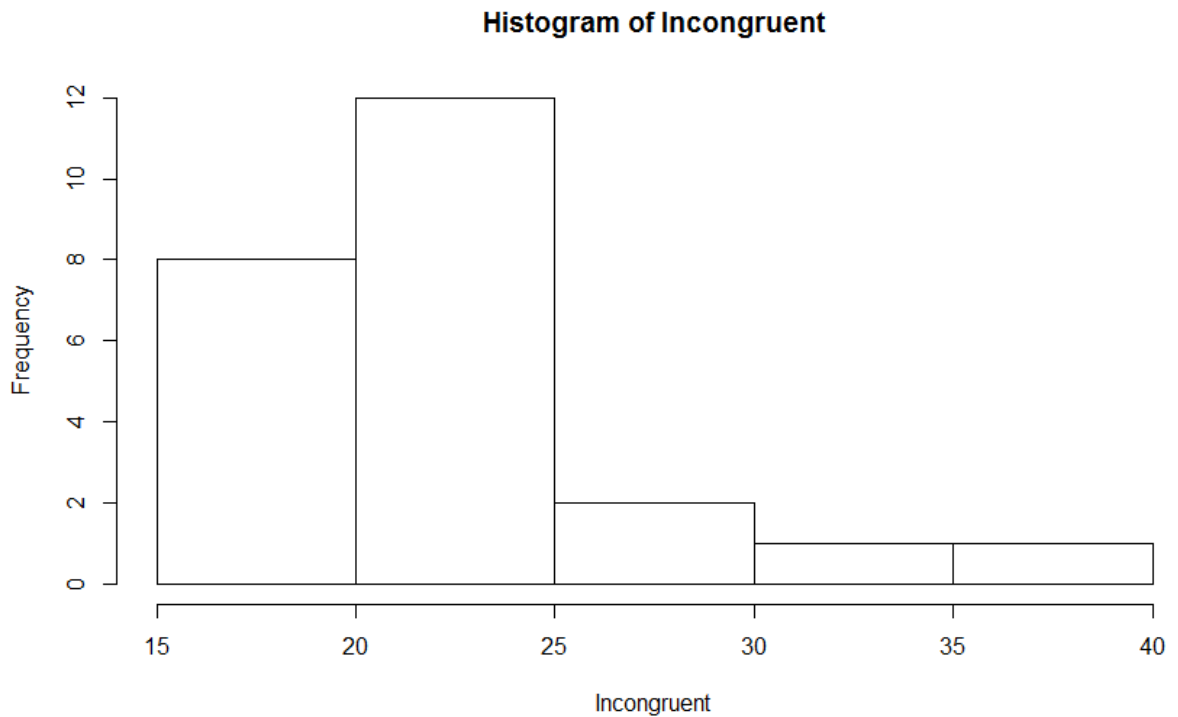
with positive skewness. With the exception of some outliers, it seems fair enough to assume the sample data is from a population of normal distribution.

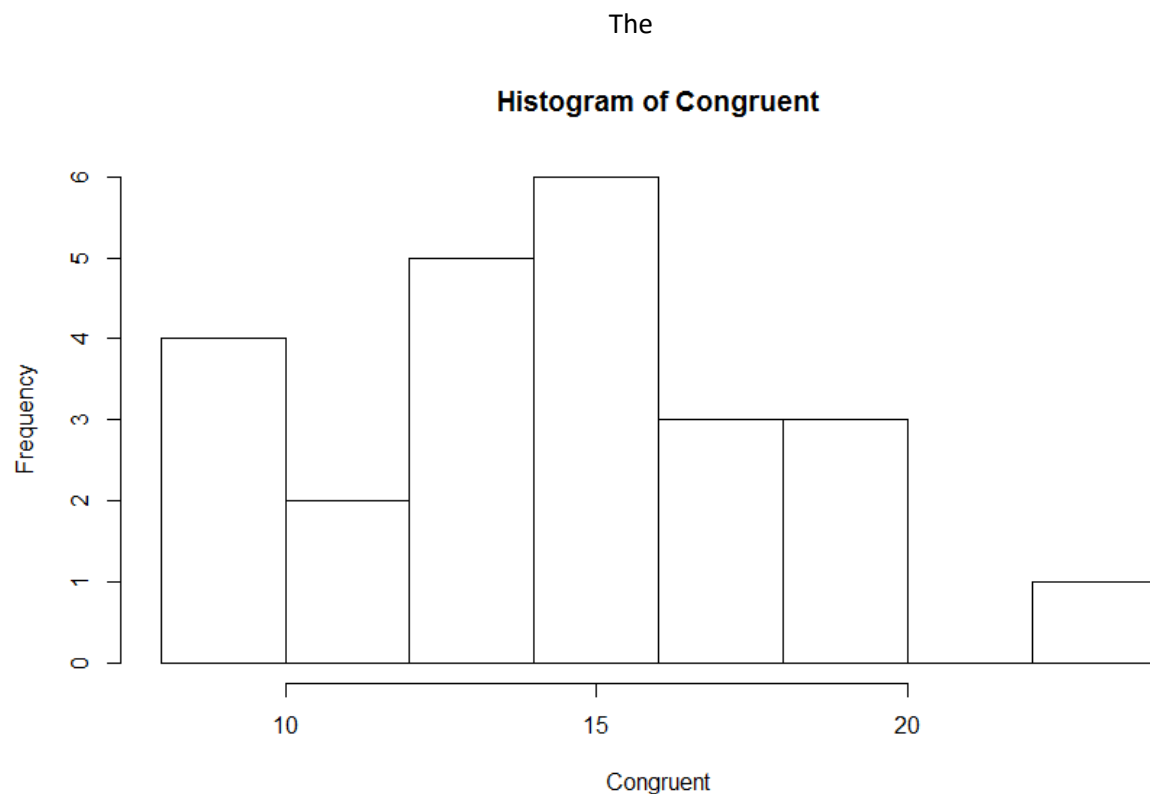
3. Report some descriptive statistics regarding this dataset. Include at least one measure of central tendency and at least one measure of variability.

[Summary \(stroopdata\)](#)

	Congruent	Incongruent	Difference
Min.	8.63	Min. :15.69	Min. : 1.950
1st Qu.	11.90	1st Qu.:18.72	1st Qu.: 3.646
Median	14.36	Median :21.02	Median : 7.667
Mean	14.05	Mean :22.02	Mean : 7.965
3rd Qu.	16.20	3rd Qu.:24.05	3rd Qu.:10.258
Max.	22.33	Max. :35.26	Max. :21.919
Var.	12.66903	23.01176	23.66654

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 histogram for both the congruent and incongruent reading time data shows normal distribution pattern with very few outliers. Even though, our sample size is small, we can still assume that, the sample reading time are from a population with normal distribution.

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?

Reviewed my test statistic and here is the new one performed with R

Paired t-test

```
data: Incongruent and Congruent
t = 8.0207, df = 23, p-value = 2.052e-08
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
```

```
6.262868      Inf
sample estimates:
mean of the differences
7.964792
```

Or using the t-test for the difference in reading time column:
One Sample t-test

```
data: Difference
t = 8.0207, df = 23, p-value = 2.052e-08
alternative hypothesis: true mean is greater than 0
95 percent confidence interval:
 6.262868      Inf
sample estimates:
mean of x
 7.964792
```

Using R- our paired t-test results shows: t-Statistic=8.02, df=23 (n-1), p-value = 2.052e-08 (which is almost 0) at 95% confidence interval. Since the p-value is less than 0.05 ($P\text{-value} < \alpha$), we reject the null hypothesis. That means there is a significant difference in the mean reading time of the two set of words. (The mean reading time of incongruent words is significantly greater than that of congruent words). Our result matches with our expectations of reading time is more among the incongruent words than congruent set of words.

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!

It's a learned and adopted learning habits. We get accustomed to one kind of method and if there is any slight change to what we get used to, it affects our efficiency. For example typing speed on two different keyboards (QWERTY and ABC). I am accustomed to the QWERTY kind of keyboard and I am sure, I would be a much slow typist for any slight change to the keyboard (change in position of the letters and key functions).

Citation

1. <http://www.ncl.ac.uk/itservice/dataanalysis/simpletests/ttests/pairedsamplesttestdependentsamplesttest/>
2. <https://cyfar.org/types-statistical-tests>
3. <https://www.youtube.com/playlist?list=PLqzoL9-eJTNBDdKgJgJzaQcY6OXmsXAHU>
4. <http://www.shodor.org/interactivate/activities/Histogram/>
5. http://www.ncss.com/wp-content/themes/ncss/pdf/Procedures/NCSS/Paired_T-Test.pdf