The Stroop Effect: Some Statistical Inferences

Shivendra Sharma

17/01/2018

Contents

Introduction to The Stroop Effect

The Stroop Effect is a popular psychological test that is used in clinical practices and investigation. It was first coined by Dr. John Ridley Stroop who first published its paper in 1935 but was earlier published in Germany in 1929.

The Stroop Effect involves reading a passage that includes several colours with their names written down. The test subject is required to note down the time in seconds or minutes taken to finish reading the passage. This is usually the first phase of the experiment. In the second phase, the test subject is required to read the ink colour of a word rather than the noun it writes down. For example, the word RED would be written in colour blue, so the test subject is required to read 'BLUE' rather than RED which is what is first observed. The subject is required to note down the time taken in reading words of phase I and the correct ink colours in phase II and calculate the difference between the two. People with lower differences in response times receive higher scores.

Through this experiment, psychologists are able to determine the processing speed and selective attention abilities of an individual and recommend clinical treatments thereon. It also tests for other cognitive abilities as well, such as parallel distributed processing. Details can be found all over the internet.

In this report, we research on the relationship between congruent words and response times in second phase, controlling words in phase II and study whether there is a significant difference between the repose timings in the two phases. We achieve this by establishing a statistical test, noting down its assumptions and applying it on our data for conclusions.

Question 1: Recognising Dependent and Independent Variables

In statistical research, one of the primary goals is establishing a relationship between what statisticians call 'dependent' variables and 'independent' variables. Dependent variables are also called as the explanatory or the predictor variable that predict the 'outcome' variable. For example, studying the relationship between industrial pollution and the number of instances of lung related disorders among a plant's workers involves the predictor variable recognised here as 'industrial pollution' and the degree of lung infections as the outcome or the predicted variable. Such a study in turn will involve several similar predictors and multiple outcomes as well. Or it can also have multiple predictors for a single outcome.

In simplistic terms, the concept known as regression analysis is mathematically written down as: $Y = \alpha + \beta X + \epsilon$
Where, α is the intercept of a regression line that visualizes how the response variable Y is behaving according

to the predictor *X*. The β coefficient is the **slope** coefficient and measures how steep the regression line is. Its always a number between 0 and 1 with 0 implying no slope at all (a horizontal line) and 1 implies a perfectly vertical line.

In Stroop Effect, there are two phases. In phase I, the test subject is allowed to read a section of words that are coloured in the ink they read. For example, the word RED will be coloured in red colour. The first phase is usually taken as a light phase and the subject is instructed to read on the full section and note down the time taken by him/her to finish that phase. In phase II, the experiment gets harder and the test subject is now instructed to read the specific ink of a word, and not the word itself. For example, the word RED might be written in blue ink; so the subject is required to read BLUE instead of RED which is what is observed at first sight.

In this manner, the individual with faster reading times scores higher on the abilities the experiment tests and thus records lower time lags in finishing phase II. In our research, we study whether an individual scoring high in phase I scores similar in phase II as well.

Let's now load our data to have a look of its structure and decide on the variables that are to be assigned.

```
stroop <- read_csv('stroopdata.csv')
```

```
## Parsed with column specification:
## cols(
##   Congruent = col_double(),
##   Incongruent = col_double()
## )
```

```
stroop
```

```
## # A tibble: 24 x 2
##    Congruent Incongruent
##        <dbl>       <dbl>
## 1     12.079      19.278
## 2     16.791      18.741
## 3      9.564      21.214
## 4      8.630      15.687
## 5     14.669      22.803
## 6     12.238      20.878
## 7     14.692      24.572
## 8      8.987      17.394
## 9      9.401      20.762
## 10    14.480      26.282
## # ... with 14 more rows
```

So, there are two variables here: `Congruent` and `Incongruent`. `Congruent` refers to phase I of the experiment where the subjects read on the section in a normal manner. As we observe, phase I has lower times recorded because it is easier and effortless to read simple words. However, phase II `Incongruent` variable records much higher times since this phase tests the abilities of a subject in full swing. Since its always harder to read the ink of a word rather than the noun, added to it the pressure of finishing the section in as little time possible, the numbers (seconds) are thus higher.

In our report, we're interested in the effect word congruency has on incongruency and the time taken for a test subject to read the ink colours of all the words correctly in phase II. Since we're taking control of the congruent condition and studying its effect on the latter phase, we therefore set `Congruent` as the predictor variable here and `Incongruent` as the dependent or outcome variable.

Now, before proceeding with our analysis, its important to note that our original data contains certain outliers that need to be dealt with. A close observation will tell us that rows 15 and 20 of the above dataset are acting as outliers. Let's remove them and make a filtered data so we'll have an original dataset and a filtered

dataset for clearer comparisons.

```
new.stroop <- stroop[-c(15, 20), ]
```

Question 2a: Establishing Appropriate Hypothesis

Hypothetical testing is a process in statistics where a researcher tests a claim through several mathematical formulae (also known as statistical tests) and measures and comments on the results extracted. One of the most popular statistical tests in testing hypothesis are the Student's t-test and a z-test.

In statistical hypothesis, the researcher formulates a null hypothesis, mathematically denoted as H_0 and the alternative hypothesis denoted as H_a. While the null hypothesis usually states something neutral for a concerned data, the alternative hypothesis states something opposite the null hypothesis. For example, referring to one of our pollution related instances above, a researcher could claim that provided a statistic for a particular pollutant, there can either be a significant difference in the instances of lung infections observed vs. no significant difference in the instances. The latter is thus a null hypothesis while the former is its alternative. There can be numerous similar examples to think of.

In these experiments, a statistical measure is picked up and tested through one of the many available statistical tests. This provides the researcher with a new statistic that he/she compares with a statistical table after selecting a significance level (also called the *rejection criteria*). The significance level can be thought of a threshold wherein if a statistic being tested lies within those limits of threshold, then the null hypothesis is accepted and vice versa. Hence, hypothesis testing is a very effective way to test the authenticity of a claim through statistical evidences.

In our case, we're required to test the mean difference in response times between congruent and incongruent conditions. Let's denote mean timings in congruent conditions as μ_c and mean timings for incongruent conditions as μ_i. Our null hypothesis is that there is no significant difference between mean timings of the two conditions vs. the alternative claim that timings in incongruent conditions is significantly larger.

Mathematically, we can write this down as:

H_0 : mean timings of the two conditions are equal, $\mu_c = \mu_i$.

H_a : mean timing of incongruent condition is larger than congruent condition, $\mu_c < \mu_i$

Question 2b: Establishing a Statistical Test

For small samples with $n \leq 30$, there are two tests namely the z-test and the Student's t-test. For most of the part, their assumptions are similar and hence their applications too. The most important of these is the assumption of normality which states that the samples that have been derived from larger populations are themselves normally distributed. Z-test and the t-test thus follow the same rule of normality.

Despite many similarities, these tests also differ when it comes to known and unknown values. The z-test makes use of the population standard deviation σ and assumes it to be known whereas the t-test assumes that the standard deviation is unknown for that population and hence uses an estimated standard deviation s Since latter is the case for most real-life distributions, researchers use t-test more with small samples. Mathematically, these tests are expressed as:

$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$

$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$

Notice the similarity in their expressions. Only their usage of σ and s differ.

The Student's t-test was invented by William Gosset who worked with the Guinness brewery in Dublin, Ireland. He introduced the t-test as a simple and affordable way of quality test in his brewery and published

his findings in 1908. Since the brewery prohibited its chemists on publishing their findings with their proper names, William thus published his through the pseudonym 'Student'. So, when should the z-test and the t-test be used?

There is a twist to the application of these statistical tests. While large sample tests can be used for small samples, the other way round is not possible and hence small sample tests are impossible to use with large samples. Specifically, the z-test is recommended when $n > 30$ and the population standard deviation is known. Else, when observations are less than 30 as has been stated already, we use the t-test with an unknown standard deviation (which is approximated and denoted as s).

Let's list down some assumptions of our data: - The samples are normally distributed. - There is little skewness in the data that would divert calculated values. - Standard deviations remain unknown. - Population variances are equal. - Samples have been derived through random sampling from a population and there is no bias. - Variable means are dependent on each other.

There are paired and unpaired t-tests as well. The former uses a control to effectively influence another event while unpaired tests do not use such influences instead using data independently. On light of such assumptions, we will thus use a paired t-test for independent means.

Also, since the original data contains outliers, it automatically violates some of the assumptions of the t-test, so Ill go ahead with the filtered data. As we shall see later one, the skewness that is observed in the original data is eliminated in the filtered data thus giving us a correct values to conduct an appropriate test and concluding without commenting any errors.

Question 3: Reporting Descriptive Statistics

One of the best ways in R to summarise data is to call the `summary()` function. This little function lists out almost every important statistic that is required. Also, since our datasets do not suffer from NA values, accurate values will this be presented.

```
summary(stroop)
```

```
##    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
```

```
summary(new.stroop)
```

```
##    Congruent        Incongruent
##  Min.   : 8.63   Min.   :15.69
##  1st Qu.:11.53   1st Qu.:18.67
##  Median :14.36   Median :20.82
##  Mean   :13.94   Mean   :20.86
##  3rd Qu.:15.83   3rd Qu.:22.64
##  Max.   :22.33   Max.   :26.28
```

While the first set of summaries refers to the original `stroop` data, the second set refers to `new.stroop`. For testing the Gaussian criterion of the two datasets, we can call the `sd()` and `var()` functions respectively to get an idea of how normally distributed the variables are.

```
print(c('The std. deviation for the original congruency is', sd(stroop$Congruent)))
```

```
## [1] "The std. deviation for the original congruency is"
## [2] "3.5593579576452"
```

```r
print(c('The variance for the original congruency is', var(stroop$Congruent)))
```

```
## [1] "The variance for the original congruency is"
## [2] "12.6690290706522"
```

```r
print(c('The std. deviation for filtered congruency is', sd(new.stroop$Congruent)))
```

```
## [1] "The std. deviation for filtered congruency is"
## [2] "3.59277301047003"
```

```r
print(c('The variance for filtered congruency is', var(new.stroop$Congruent)))
```

```
## [1] "The variance for filtered congruency is"
## [2] "12.9080179047619"
```

And there we have the details regarding the `Congruent` variable for the original and filtered data. We print similar results for the `Incongruent` variable.

```r
print(c('The std. deviation for the original incongruency is', sd(stroop$Incongruent)))
```

```
## [1] "The std. deviation for the original incongruency is"
## [2] "4.79705712246914"
```

```r
print(c('The variance for the original incongruency is', var(stroop$Incongruent)))
```

```
## [1] "The variance for the original incongruency is"
## [2] "23.0117570362319"
```

```r
print(c('The std. deviation for filtered incongruency is', sd(new.stroop$Incongruent)))
```

```
## [1] "The std. deviation for filtered incongruency is"
## [2] "2.8769225023305"
```

```r
print(c('The variance for filtered incongruency is', var(new.stroop$Incongruent)))
```

```
## [1] "The variance for filtered incongruency is"
## [2] "8.27668308441558"
```

So, while the statistics remain pretty much similar for `Congruent`, there are differences for `Incongruent` under the filtered data. The second set of data certainly seems more Gaussian. Visualizations for these distributions will make the scenario even clearer regarding the normality assumption. There are additional conclusions that we come across regarding the filtered data. The assumption of equal variances is closer in the case of filtered data while the values differ by a vast difference in the original data. Hence, our assumptions are more in concord with the the `new.stroop` dataset.

Question 4: Visualizating the Spread of Data

The project requires us to show some basic visualizations and comment on the spread of the dataset. Let's present a simple histogram of the `Congruent` data and then for the `Incongruent` variable.

```r
histo <- ggplot(stroop, aes(Congruent))
histo + geom_histogram(bins = 11, binwidth = 2, color = 'blue', fill = 'red') + ggtitle(label = 'Histog
```

Histogram of Original Congruency
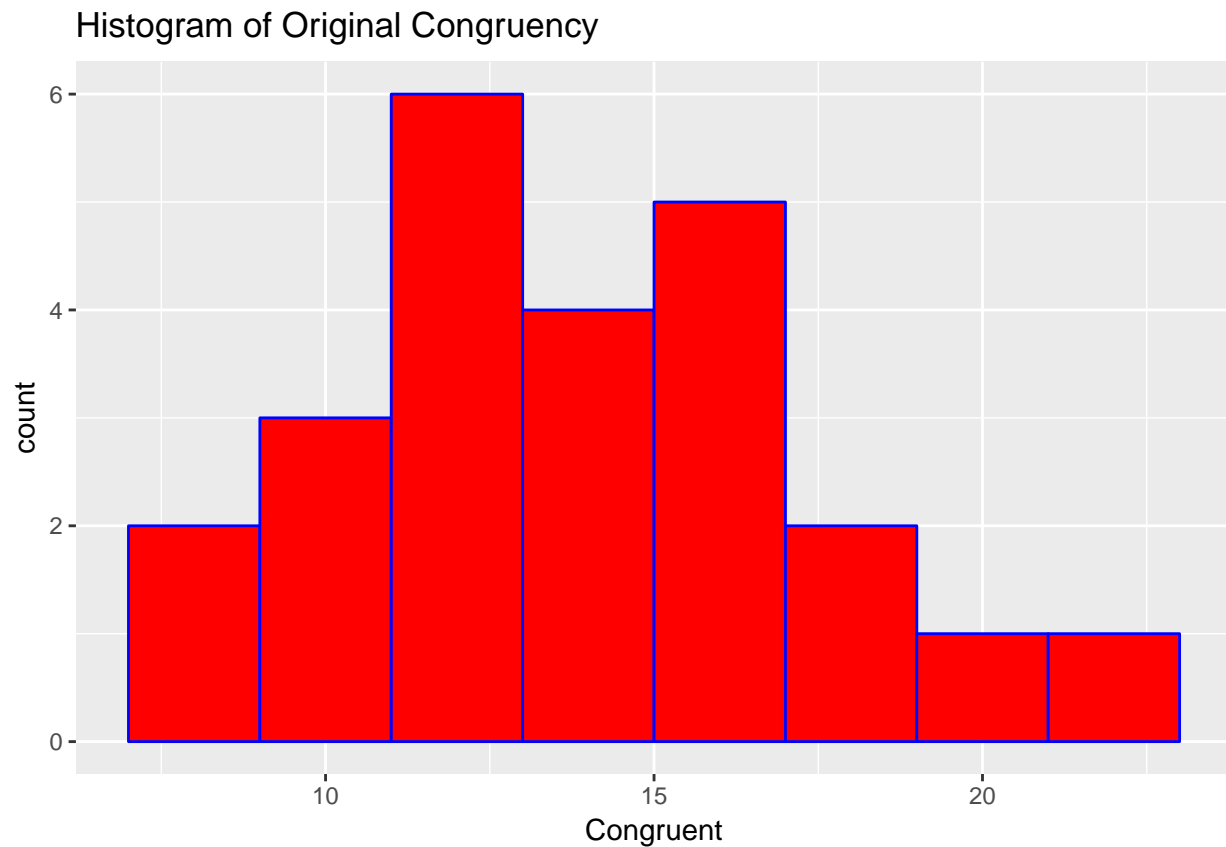
So the `Congruent` variable seems normal in the original data. Let's reproduce the `Incongruent` variable.
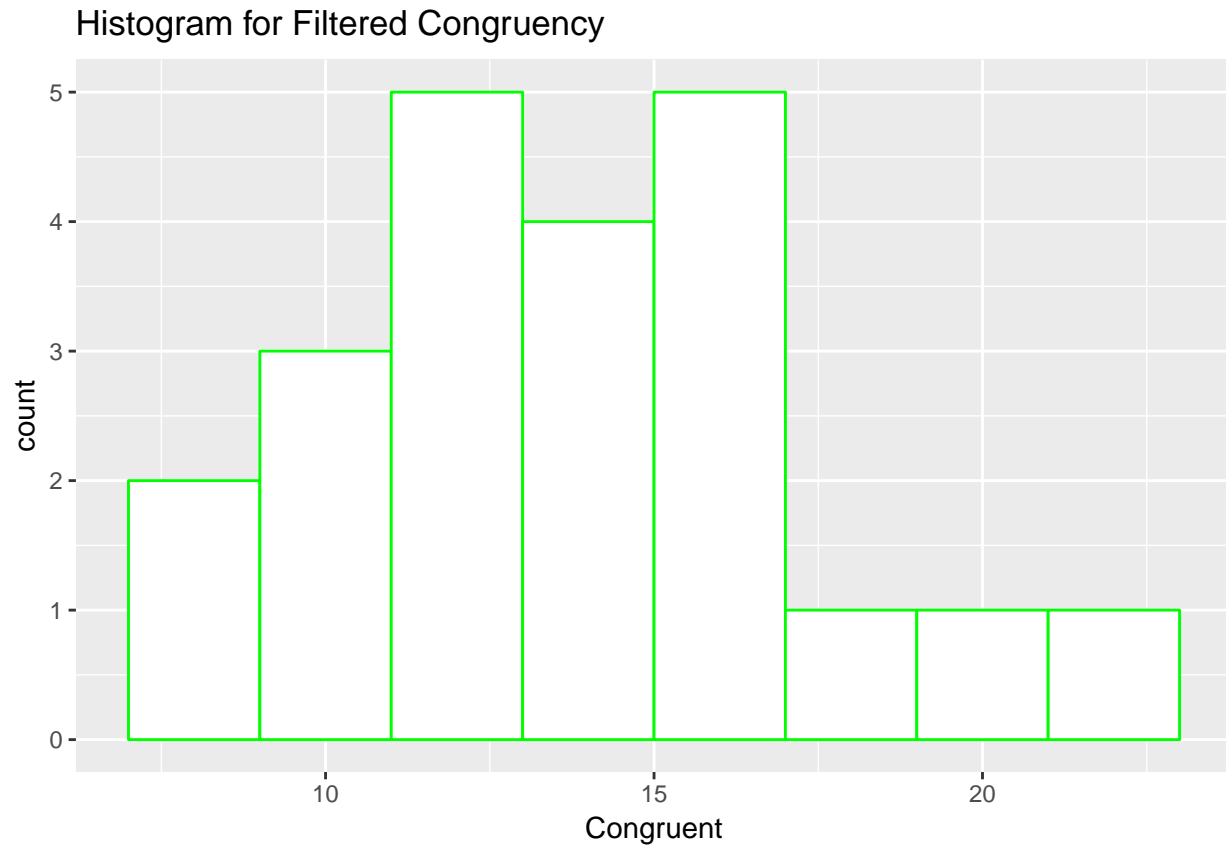
```
histo <- ggplot(stroop, aes(x = Incongruent))
histo + geom_histogram(bins = 11, color = 'blue', fill = 'red', binwidth = 2) + ggtitle('Histogram for
```

Histogram for Original Incongruency
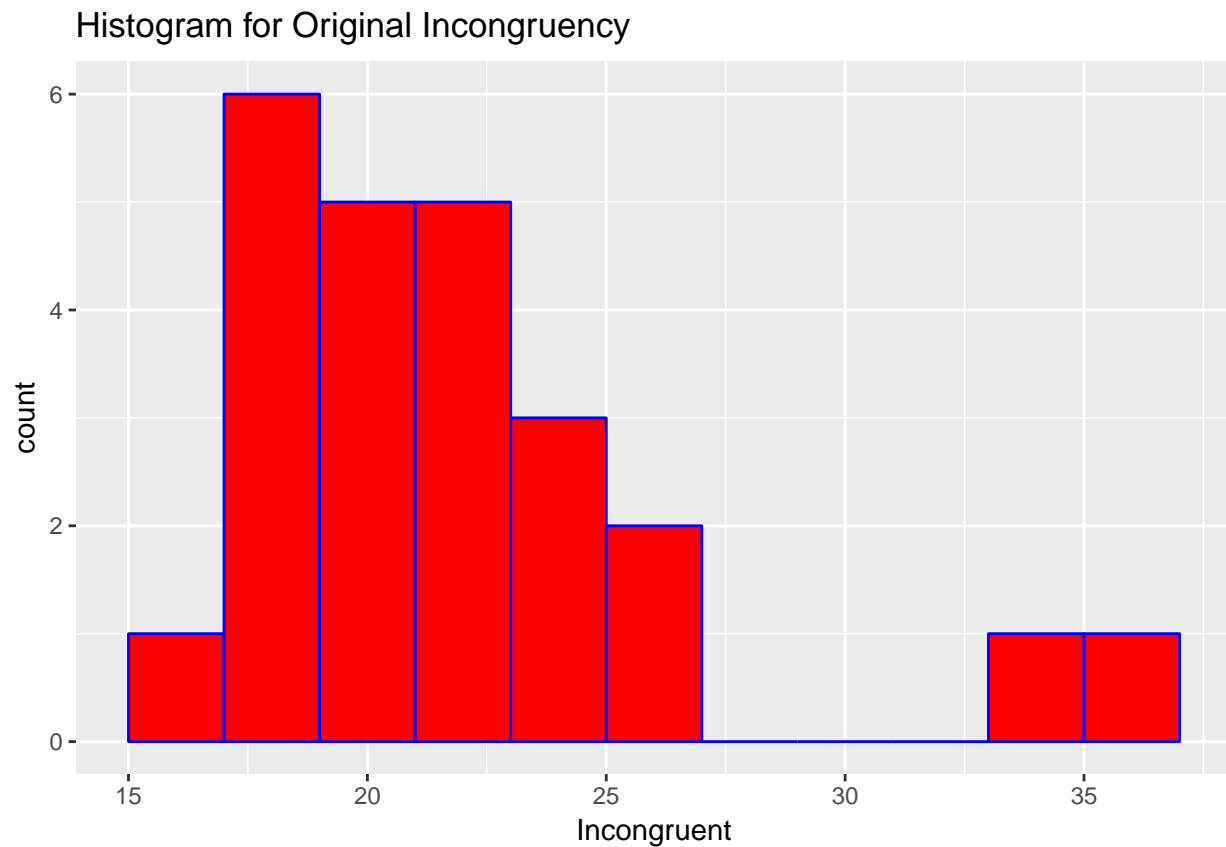
That's a major diversion from the normality assumption and more inclined towards a positively skewed distribution. Those two bars on the far right are clearly the outlier observations.

Let's compare now for the filtered data.

```
histo <- ggplot(new.stroop, aes(x = Congruent))
histo + geom_histogram(bins = 11, color = 'green', fill = 'white', binwidth = 2) +

        ggtitle('Histogram for Filtered Congruency')
```

Histogram for Filtered Congruency

Comparing this with the original distribution, I feel the filtered data for `Congruent` is more skewed. Similarly, for the `Incongruent` variable, we have the following histogram.

```
histo <- ggplot(new.stroop, aes(x = Incongruent))
histo + geom_histogram(bins = 11, color = 'green', fill = 'white', binwidth = 2) +

        ggtitle('Histogram for Filtered Incongruency')
```

Histogram for Filtered Incongruency
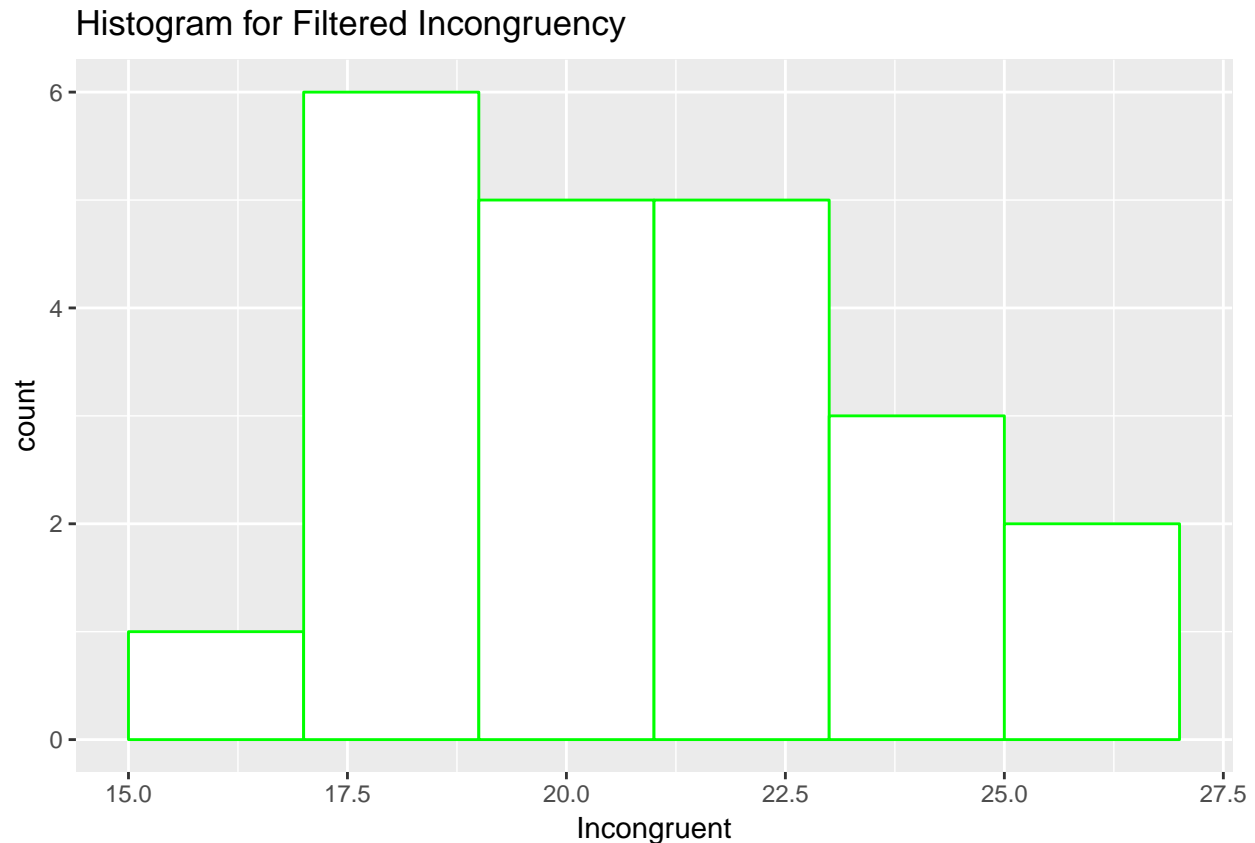
An incomparable difference! While the original variable was a skewed distribution, the filtered variable looks very normal. One reason why removing the outliers is one of the most essential tasks of efficient data analysis.

Question 5: Testing the Hypothesis

As mentioned above, we'll be using the paired t-test for paired data. Reiterating this mathematically, the t-test is expressed as $t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$, where $\bar{x}$ is the real mean, μ_0 is the hypothetical mean, s is the approximated standard deviation and n denotes number of observations. Calling a `t.test` with `paired = TRUE`, and the default 95% confidence intervals, we thus get the following values:

```
t.test(new.stroop$Congruent, new.stroop$Incongruent, paired = TRUE)
```

```
##
##  Paired t-test
##
## data:  new.stroop$Congruent and new.stroop$Incongruent
## t = -9.5411, df = 21, p-value = 4.382e-09
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -8.425043 -5.409594
## sample estimates:
## mean of the differences
##               -6.917318
```

And so, there we have the paired results for the `new.stroop` data. Important numbers to be noted here are the p-value (4.382e-09 ~ 0.0000000004382), the t-value (-9.5411), the 95% confidence interval values and the mean of the differences (-6.9173).

The most important here is the t-value of -9.5411. Comparing this with the numerals for significance levels, we conclude that mean time responses for `Congruent` and `Incongruent` are significant since the t-value is well outside the left most limit of -8.4525 is well under the calculated t-value. Notice also the p-value which is 4.382×10^{-9} (~4.382e-09). There are usually three interpretations for the p-values: - A value ≤ 0.05 primarily indicates strong evidence against the null hypothesis. - A value $= 0.05$ indicates the test can go either way. - A value ≥ 0.05 indicating evidence against the alternative hypothesis.

Since our p-value is extremely small and less than 0.05, we thus have strong evidence that the null hypothesis is rejected. There is thus a significant difference between average response times under congruent and incongruent conditions.

Question 6: Digging Deeper and Extending Investigation

With our t and p-values providing us with everything to support one or the other hypothesis, most of our exploratory and testing work is now complete. Let's ask the question of what could be the reasons for the marginal acceptance of the alternative hypothesis since we're going more with the filtered data.

In the Stroop Effect, psychologists are interested in testing out selective attention and thereby the processing speed of a test subject. In my personal experience, congruent words only marginally tested my attention and processing speed that were otherwise tested considerably in the second part of the experiment. So the experiment was one-sided. Our results too tell us a similar story. `Incongruent` variable is only marginally dependent on how one scores in `Congruent`.

We can think of similar such experiments. For example, we can think of researching on mathematical skills vs. programming skills. Although both are scientific subjects, it is not necessary that being good in one will decide being good at another. Both subjects require separate practise and experience for becoming 'fluent' although there can be a minor relationship between being inherently good at maths and being somewhat better at programming compared to one's peers.

This concludes my project for the second phase of our program. Hoping that it was informative!