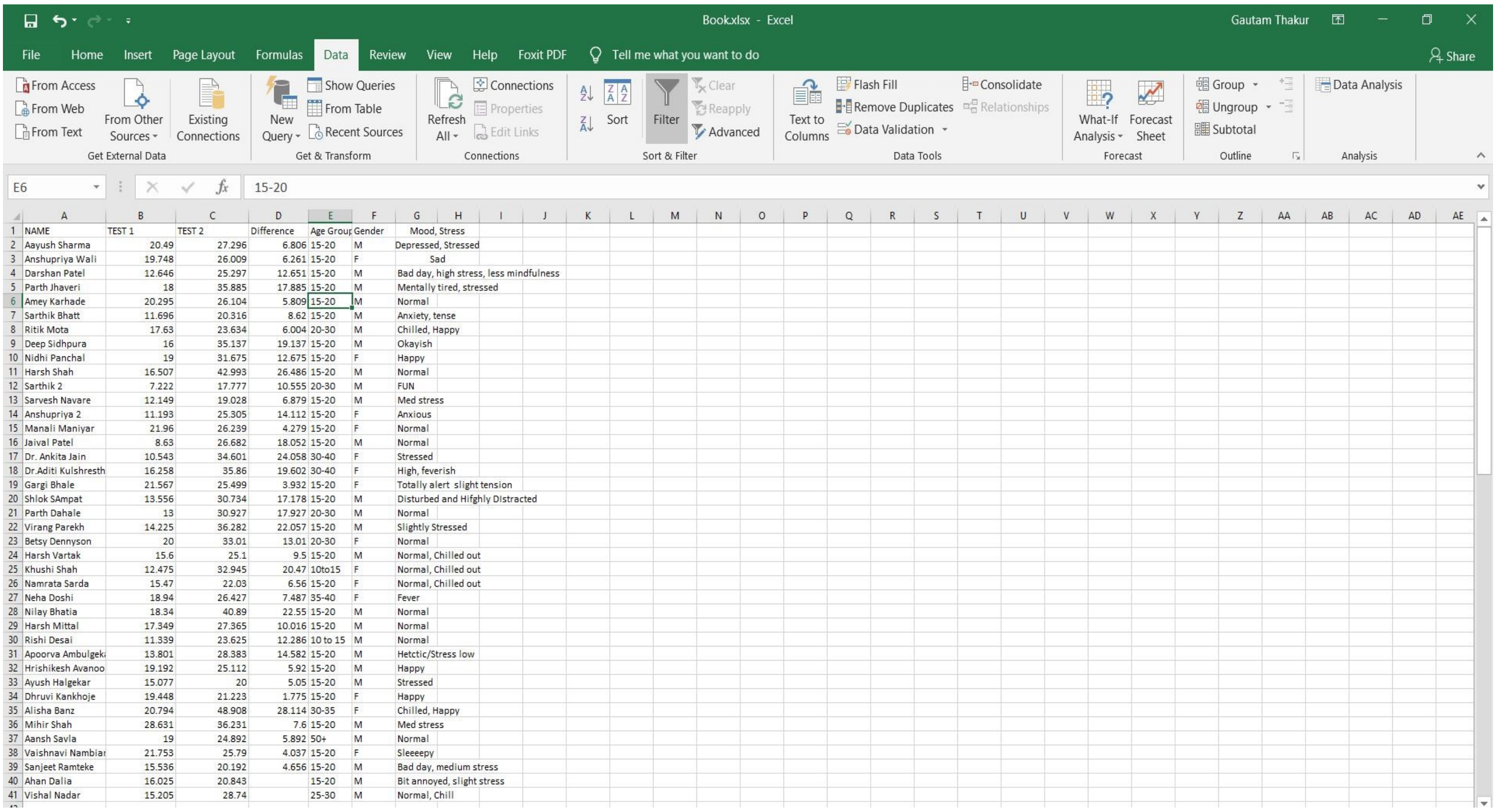
UDACITY PROJECT: STROOP TEST

As a student and a fresher at such a project, I was quite impressed with the Stroop Effect, observing how different people react to the changed conditions. Attached with this document is a copy of the raw data I have generated myself, conducting these tests with around 40 people. I tried my best to get varying kinds of inputs given from different people, due to several constructs such as Age, Gender, Preparedness, Stress levels etc.



	NAME	TEST 1	TEST 2	Difference	Age Group	Gender	Mood, Stress
2	Aayush Sharma	20.49	27.296	6.806	15-20	M	Depressed, Stressed
3	Anshupriya Wali	19.748	26.009	6.261	15-20	F	Sad
4	Darshan Patel	12.646	25.297	12.651	15-20	M	Bad day, high stress, less mindfulness
5	Parth Jhaveri	18	35.885	17.885	15-20	M	Mentally tired, stressed
6	Amey Karhade	20.295	26.104	5.809	15-20	M	Normal
7	Sarthik Bhatt	11.696	20.316	8.62	15-20	M	Anxiety, tense
8	Ritik Mota	17.63	23.634	6.004	20-30	M	Chilled, Happy
9	Deep Sidhpura	16	35.137	19.137	15-20	M	Okayish
10	Nidhi Panchal	19	31.675	12.675	15-20	F	Happy
11	Harsh Shah	16.507	42.993	26.486	15-20	M	Normal
12	Sarthik 2	7.222	17.777	10.555	20-30	M	FUN
13	Sarvesh Navare	12.149	19.028	6.879	15-20	M	Med stress
14	Anshupriya 2	11.193	25.305	14.112	15-20	F	Anxious
15	Manali Maniyar	21.96	26.239	4.279	15-20	F	Normal
16	Jaival Patel	8.63	26.682	18.052	15-20	M	Normal
17	Dr. Ankita Jain	10.543	34.601	24.058	30-40	F	Stressed
18	Dr.Aditi Kulshresth	16.258	35.86	19.602	30-40	F	High, feverish
19	Gargi Bhale	21.567	25.499	3.932	15-20	F	Totally alert slight tension
20	Shlok SAmpat	13.556	30.734	17.178	15-20	M	Disturbed and Hifghly DIstracted
21	Parth Dahale	13	30.927	17.927	20-30	M	Normal
22	Virang Parekh	14.225	36.282	22.057	15-20	M	Slightly Stressed
23	Betsy Dennyson	20	33.01	13.01	20-30	F	Normal
24	Harsh Vartak	15.6	25.1	9.5	15-20	M	Normal, Chilled out
25	Khushi Shah	12.475	32.945	20.47	10to15	F	Normal, Chilled out
26	Namrata Sarda	15.47	22.03	6.56	15-20	F	Normal, Chilled out
27	Neha Doshi	18.94	26.427	7.487	35-40	F	Fever
28	Nilay Bhatia	18.34	40.89	22.55	15-20	M	Normal
29	Harsh Mittal	17.349	27.365	10.016	15-20	M	Normal
30	Rishi Desai	11.339	23.625	12.286	10 to 15	M	Normal
31	Apoorva Ambulgek	13.801	28.383	14.582	15-20	M	Hetctic/Stress low
32	Hrishikesh Avanoo	19.192	25.112	5.92	15-20	M	Happy
33	Ayush Halgekar	15.077	20	5.05	15-20	M	Stressed
34	Dhruvi Kankhoje	19.448	21.223	1.775	15-20	F	Happy
35	Alisha Banz	20.794	48.908	28.114	30-35	F	Chilled, Happy
36	Mihir Shah	28.631	36.231	7.6	15-20	M	Med stress
37	Aansh Savla	19	24.892	5.892	50+	M	Normal
38	Vaishnavi Nambiar	21.753	25.79	4.037	15-20	F	Sleeeepy
39	Sanjeet Ramteke	15.536	20.192	4.656	15-20	M	Bad day, medium stress
40	Ahan Dalia	16.025	20.843		15-20	M	Bit annoyed, slight stress
41	Vishal Nadar	15.205	28.74		25-30	M	Normal, Chill

Figure 1: Raw Data Gathered

The independent variable here is whether the colour matches with the text or not. Due to this, people will take different times reading the entire paragraph. Hence, Time is the dependent variable.

An appropriate set of hypotheses for this test would be:
H_0 = Null Hypothesis=Nearly everyone will take lesser time in the second test, when the colours do not match with the text.

H_A=Alternative Hypothesis=Nearly everyone will take more time in the second test, when the colours match with the text.
This will also prove, if the treatment(change of text and colour) had any effect at all.
Put mathematically, if μ_0 was the average time taken in the first test, and μ_A was the average time taken in the second test,

$$H_0 \rightarrow \mu_0 >= \mu_A$$
$$H_A \rightarrow \mu_0 < \mu_A$$

OR;

$$H_0 \rightarrow \mu_0 - \mu_A >=0 \ (\mu_D>=0)$$
$$H_A \rightarrow \mu_0 - \mu_A <0 \ ((\mu_D<0)$$

Following is the method employed to statistically prove if we can accept or reject the null:

1) Take out 50 samples of the raw data, one sample of which contains the average of 5 randomly selected variables.
2) Use the samples above, to find the average, standard deviation, and the selected confidence interval, of the sample.
3) Plot histograms of the sample sets obtained. We observe that the distributions are normal.
4) Find out critical values in the sample set of test 1. Compare these to values in that of the set for test 2.
5) If the mean of the second set falls within the confidence region of the first set, we accept the null. Else, we reject the null.

I made use of python, to generate the samples containing the average of the times of the 40 people. The following is a screenshot taken on the Jupyter Notebook:

```
In [ ]: import pandas as pd
        import numpy as np
        import random
        df=pd.read_csv('tests.csv')
        b=df['TEST 1']
        c=np.zeros(50)
        for i in range(50):
            sum=0
            for j in range(5):
                sum+=b[random.randrange(0,41)]/5
            c[i]=sum
        c=pd.DataFrame(c,columns=['TEST_1'])
        c.to_csv('eval1.csv')
```

Figure 2.a: Code for test 1

```
In [ ]: import pandas as pd
        import numpy as np
        import random
        df=pd.read_csv('tests.csv')
        b=df['TEST 2']
        c=np.zeros(50)
        for i in range(50):
            sum=0
            for j in range(5):
                sum+=b[random.randrange(0,41)]/5
            c[i]=sum
        c=pd.DataFrame(c,columns=['TEST_2'])
        c.to_csv('eval2.csv')
```

Figure 2.b: Code for test 2

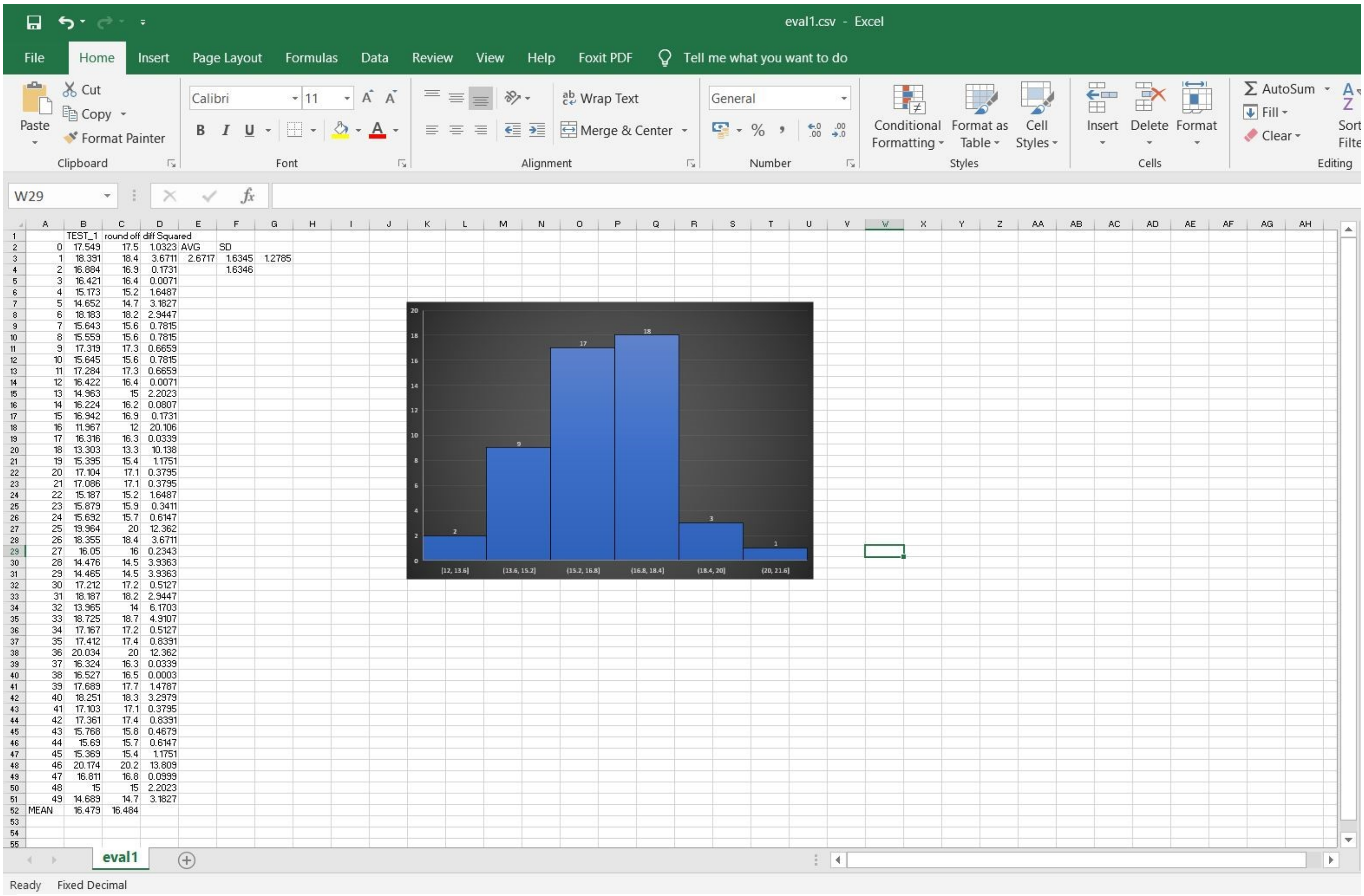


Figure 3.a: csv file for test 1

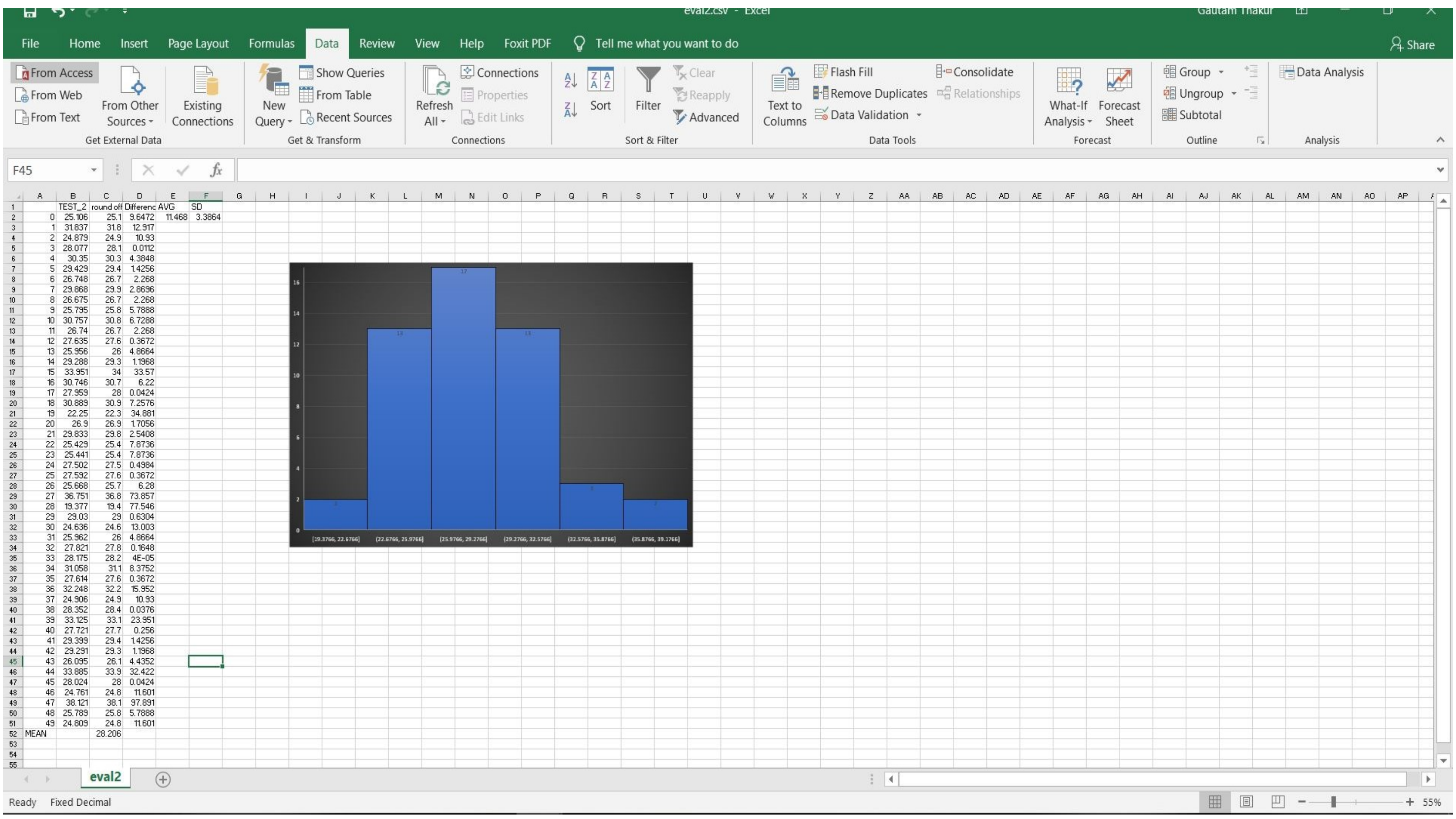
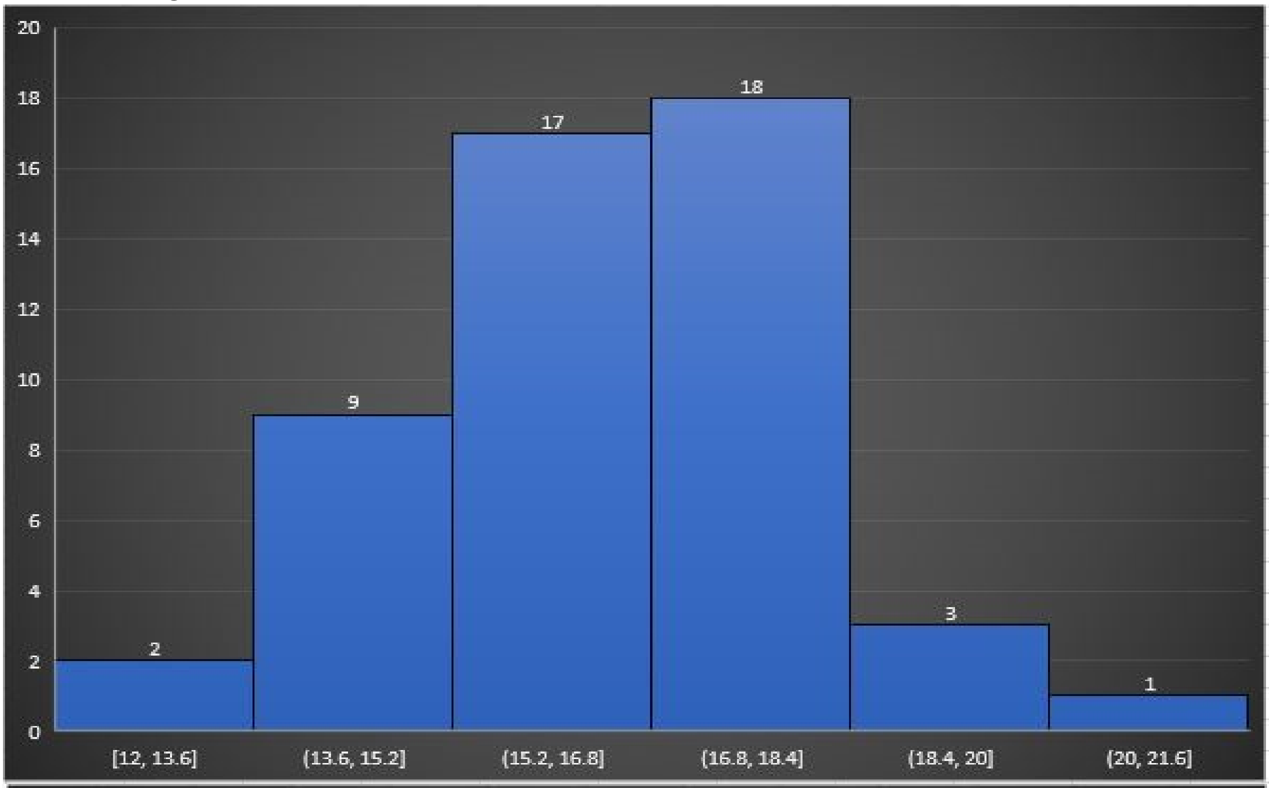
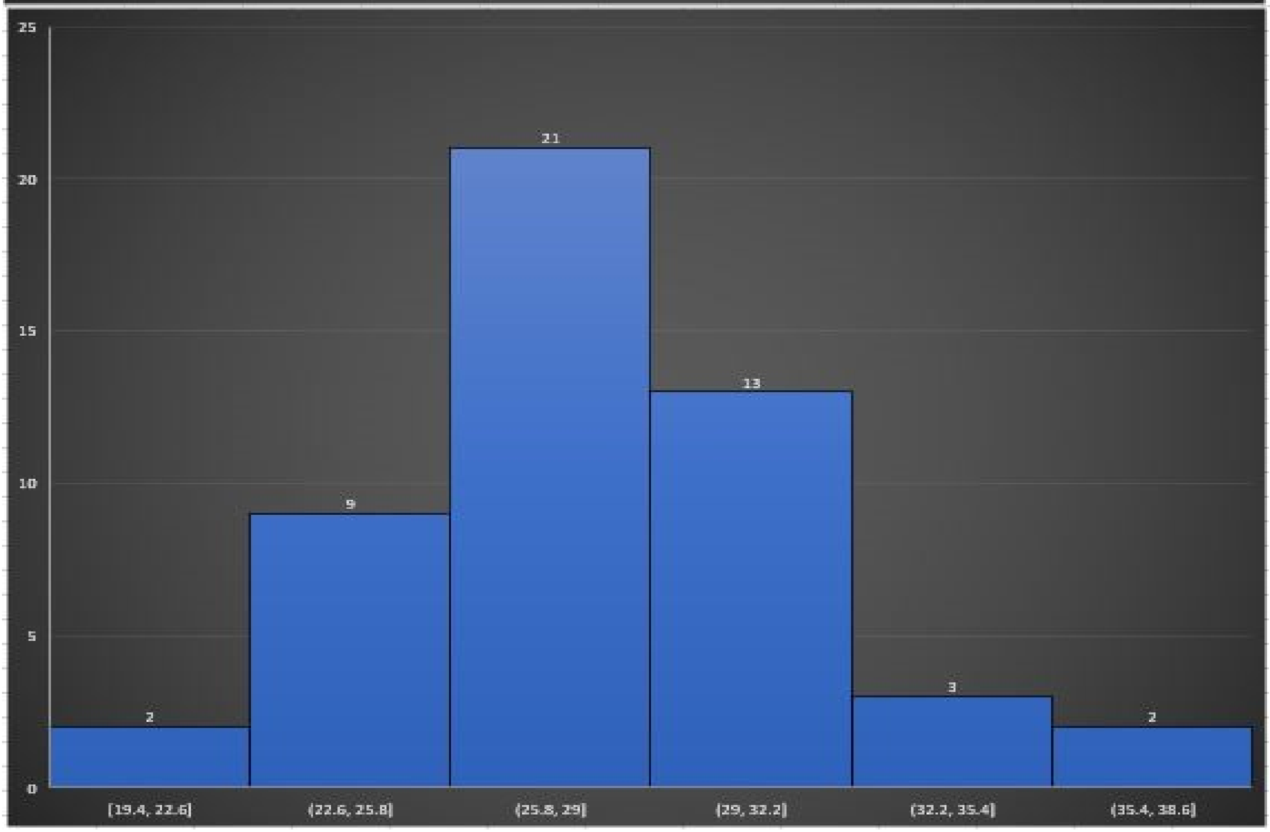


Figure 3.b: csv file for test 2

The csv (Comma-Separated-Values) files at the end of each execution of the cells contain 50 samples, each having the average of 5 randomly selected times in both test cases. In order to save complexity, I have rounded of the samples to 1 decimal value. I have made the use of Excel to calculate the average and the Standard Deviations of the samples, storing all the data in the same respective csv files. Using this data, I plotted Histograms for both csv files obtained. Following are the 2 histograms:





Sample Average for test 1 ($\overline{x1}$) = 16.484s **Standard Deviation of samples in test 1 ($\sigma1$) = 1.635s**

Sample Average for test 2 ($\overline{x2}$) = 28.206s **Standard Deviation of samples in test 2 ($\sigma 2$) = 3.386s**

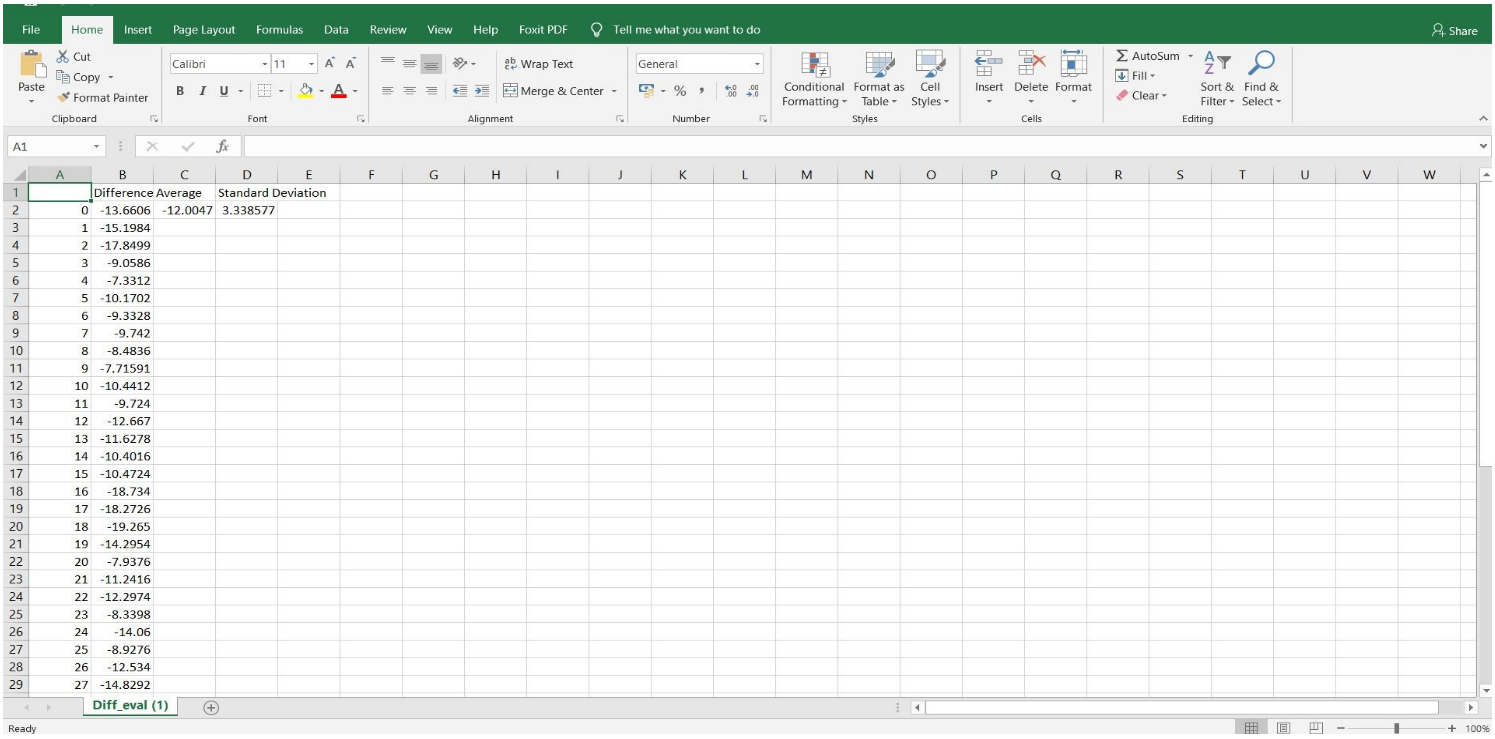
Since we are going to only compare 2 values to check which is greater, a one-tailed test should be sufficient. Therefore, my critical region is only going to be in the lower region of the plot obtained in test 2. I base my calculations on the t-tests, as we don't know the population mean, which rules out the Z tests.

I have decided to keep a Confidence Interval (CI) of **95%.**

So, if the t values obtained go into the critical region, i.e. less than alpha level 0.05, we can say that the average time taken for the tests are definitely different, and that the time taken in test 2 is greater than that in test 1(Rejecting Null). That is because we cannot obtain such a value at random, as the probability of obtaining that is below 0.05. The changed factor (colour does not match with the text) will have definitely caused a change.

The point estimate for the above study is $\overline{x1}$-$\overline{x2}$=16.484-28.206=-11.72
The sample size (n) selected is 5. That means, my degrees of freedom will be n-1=5-1=4.
Now, using the T-table, the t-critical value is 2.132.
A few more calculations are required, before we get comparing the t values. Using python, I have calculated the difference of the times in the above to samples, along with the average difference and the Sample Standard Deviation (Using Bessel's Correction).

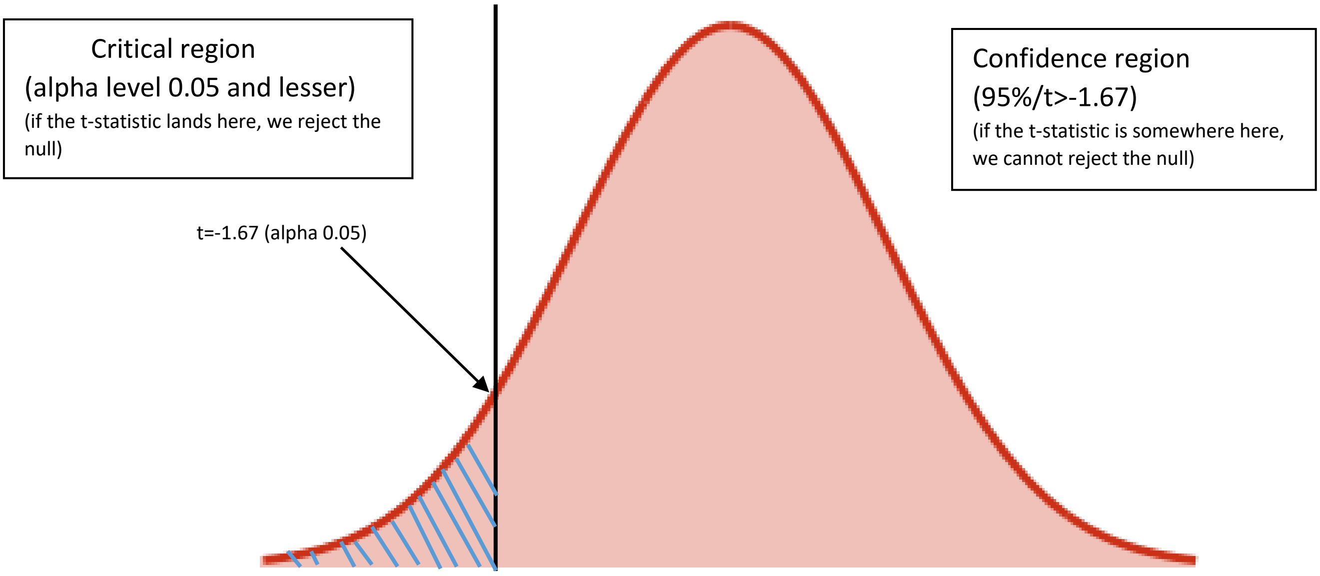
```python
import pandas as pd
import numpy as np
df1=pd.read_csv('eval1.csv')
df2=pd.read_csv('eval2.csv')
diff=np.zeros(50)
for i in range(50):
    diff[i]=df1['TEST_1'][i]-df2['TEST_2'][i]
diff=pd.DataFrame(diff,columns=['Difference'])
diff['Average']=diff['Difference'].mean()
average=diff['Difference'].mean()
sum=0
for i in range(50):
    sum+=(diff['Difference'][i]-average)**2
SD=(sum/50)**2
diff['Standard Deviation']=SD
diff['Standard Deviation'].drop_duplicates(keep='first',inplace=True)
diff['Average'].drop_duplicates(keep='first',inplace=True)
diff.to_csv('Diff_eval.csv')
```



	A	B	C	D
1		Difference	Average	Standard Deviation
2	0	-13.6606	-12.0047	3.338577
3	1	-15.1984		
4	2	-17.8499		
5	3	-9.0586		
6	4	-7.3312		
7	5	-10.1702		
8	6	-9.3328		
9	7	-9.742		
10	8	-8.4836		
11	9	-7.71591		
12	10	-10.4412		
13	11	-9.724		
14	12	-12.667		
15	13	-11.6278		
16	14	-10.4016		
17	15	-10.4724		
18	16	-18.734		
19	17	-18.2726		
20	18	-19.265		
21	19	-14.2954		
22	20	-7.9376		
23	21	-11.2416		
24	22	-12.2974		
25	23	-8.3398		
26	24	-14.06		
27	25	-8.9276		
28	26	-12.534		
29	27	-14.8292		

Diff_eval (1)

As seen, the Average Difference (μ_{diff}) is -12.0047 (Nearly equals the point estimate)
The Sample Standard Deviation (SD) is 3.3386. For better understanding, here is a visualisation of the second test's data, and the critical region we have chosen.



Finding the t-statistic of the mean difference (**$\mu_{diff=}$** -12.0047):

$$t_A = \frac{\bar{x}\mathbf{1} - \bar{x}\mathbf{2}}{SD/\sqrt{n}} = \frac{-12.0047}{3.3386/\sqrt{5}} = \mathbf{-8.0403}$$

The t-statistic is much greater than the t critical value for alpha level 0.05.
The t-critical value for an alpha level 0.01 is -3.747, and for 0.001 is -7.173

The t-statistic is still very high for the above values. This means that the probability of obtaining such a sample is lesser than 0.001, which implies that a particular treatment (changing the text) has had a notable difference in the data obtained, compared to the initial data. This is clear proof, that the time taken in the first test is neither equal to nor greater than the time in the second test.

<u>Therefore, we REJECT THE NULL.</u>

This means, that according to the above study, we can conclude that nearly everyone takes more time in the second test, than in the first. The results are as I had expected, after giving the Stroop test myself. My own score was 13.5s in the first, and 29.7s in the second.

I believe that this result is logical, as it also counts on the adaptability of the human brain. After the first test, the we subconsciously keep the mindset, that the second test too will be like the first, and focus more on speed by just reading the words themselves. We get surprised on seeing that the colours don't match, and that throws us off. However, that is only for a bit. When I conducted these tests, I didn't tell the participants that the second set will be different. In the second test, they were surprised only for a while, and nearly everyone was reading out correctly after the first 2 lines.

The difference in the time taken in the 2 tests is mostly due to the reaction of people, when their mindset/game plan fails, and how fast the brain re-adapts, focusing only on the colour of the text. That's also when the participant regains the fluency in giving this test. Granted, there are several other factors like mood, stress, age, gender, etc. but I believe the most promising answer to the result obtained was this.

Giving it some thought, I found out that similar tests could be as follows:

1) Speak a sentence out loud, letter wise, in order first. Then, say the same sentence in a reverse manner.
   ex: “I like You” will be I-l-i-k-e-y-o-u, then u-o-y-e-k-i-l-I.
2) Read out a paragraph normally at first, but then read it out in the ‘P-Language’. The P language is no Latin or Greek, but just the same old English, with a ‘P’ following every vowel of the word, followed by the vowel again.
   Ex: “Hi! My name is Katie”, will be “Hi-pi, My na-pa-me-pe i-pi-s Ka-pa-ti-pi-e-pe. The fun here is that even though like “My” sound like they may have a vowel, but they don’t!