Project 2
Udacity Machine Learning Foundations Nanodegree

Following is the work done as part of graduating for Machine-learning Foundation Nanodegree Unit – 2[Statistics]

Problem description:

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.

The data are recorded and the time taken for incongruent condition is compared with the congruent results.

Following is the report of the data from the spreadsheet at the following address:
https://www.google.com/url?q=https://drive.google.com/file/d/0B9Yf01UaIbUgQXpYb2NhZ29yX1U/view?usp%3Dsharing&sa=D&ust=1499120159475000&usg=AFQjCNFJ_LrV1gkQ1DmfvX3NHUddazc_gA

Data

Congruent	Incongruent
12.079	19.278
16.791	18.741
9.564	21.214
8.63	15.687
14.669	22.803
12.238	20.878
14.692	24.572
8.987	17.394
9.401	20.762
14.48	26.282
22.328	24.524
15.298	18.644
15.073	17.51
16.929	20.33
18.2	35.255
12.13	22.158
18.495	25.139
10.639	20.429
11.344	17.425
12.369	34.288
12.944	23.894
14.233	17.96
19.71	22.058
16.004	21.157

Data Report:
Symbols used:
      Mean of congruent samples - $\overline{X}_C$
      Mean of incongruent samples - $\overline{X}_I$
      Standard deviation – σ
      Population congruent mean – μ_C
      Population incongruent mean - μ_I

Descriptive Statistics

As per the data, the following can be inferred:

Independent variable : **Type of condition** – Congruent (word color matches with the word)/Incongruent (Word color does not match with the color)

Dependent variable   : **Time taken** by the users to complete/say out loud the words test under a given condition
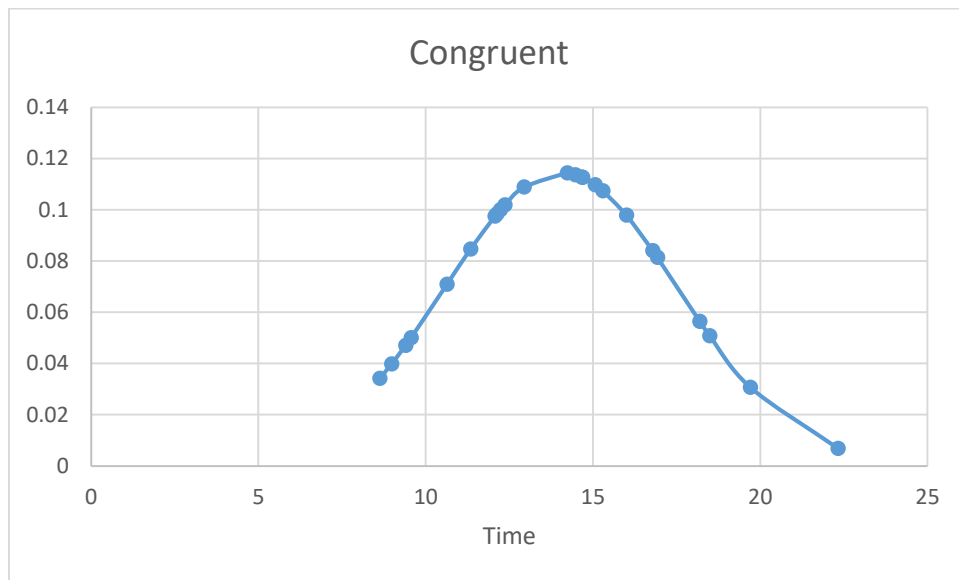
Sample size: 24

Mean congruent, $\overline{X}_C$= 14.05

Mean Incongruent, $\overline{X}_I$ = 22.01

Standard deviation, σ = 4.86

By looking at the data, the time taken to complete the incongruent test can be compared with the congruent time taken.

Sample plots:

Incongruent (chart, x-axis: Time, y-axis: 0 to 0.09)

Inferential Statistics
Defining Hypothesis:

Null hypothesis : Population Incongruent mean = Population congruent mean i.e. $\mu_I = \mu_C$

Alternative hypothesis : Population Incongruent mean > Population congruent mean i.e. $\mu_I > \mu_C$
Population Incongruent mean < Population congruent mean i.e. $\mu_I < \mu_C$
Population Incongruent mean != Population congruent mean i.e. μ_I != μ_C

Type pf test proposed
Since the sample is the same, Repeated measures test (two condition test) can be performed on the data set because the user performance is based on two different conditions (Congruent & Incongruent).
It is a one-tailed test on the negative side.
The above decision based on the assumptions/reasons below:

1. The distribution is Gaussian
2. We have less than 30 samples
3. The mean, standard deviation of the population is unknown, hence the t-test.
4. Same sample is tested at two different points of time, hence the repeated measures test.
5. Further assuming, α level=0.05

Characteristics of the data
- Mean difference$\mu = \overline{X}_C - \overline{X}_I$ = -7.96
- Degrees of freedom, df: 23
- Standard error, SEM = 0.993
- t-statistic= -8.020938234
- t-critical= -1.714

Reject null hypothesis as t-statistic is much lesser than t-critical value.
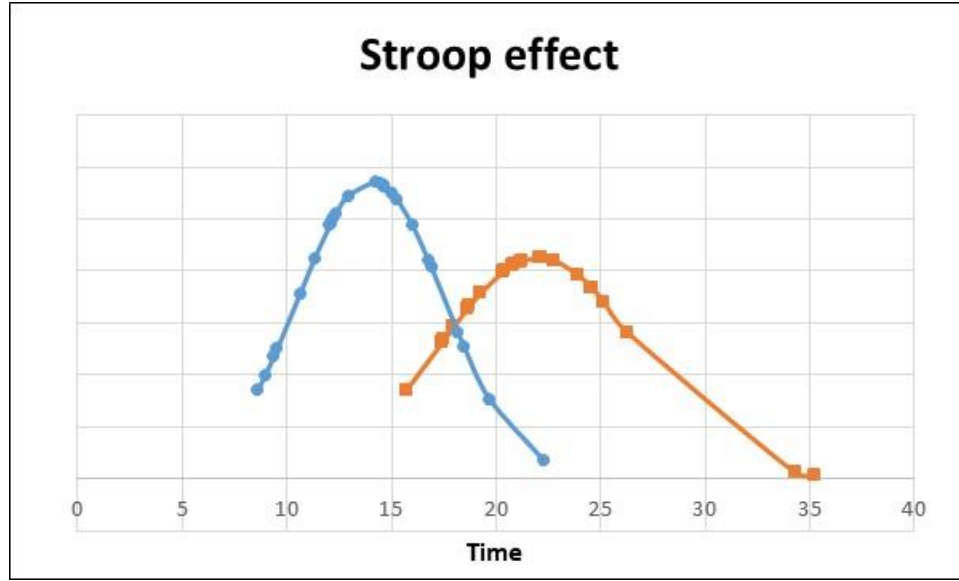Alternative hypothesis holds good for the data.

Confidence Interval, CI: (-9.67, -6.27)
Cohen's d= -1.637219949
r^2= 0.108430147

Conclusion

It can be concluded that users took 6 to 10 seconds lesser on the congruent condition.

Stroop effect

Time

Since this was an experimental test, a conclusion can be made that the incongruent condition caused an increase in time to complete the test by the users.