P1. Statistics: The Science of Decisions

1. Basic Understanding of the Stroop Effect

1.1 Introduction

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.

2. Questions:

2.1 Identification and classification of variables in the experiment

Dependent Variable:

The dependent variable was the reaction time taken to name the font colour of the text presented to the test subject.

Independent Variable:

The independent variable in this experiment was whether the word name and font color were the same or different. Another independent variable is the normal reading time of every individual which is different.

2.2 Establishment of an appropriate set of hypotheses and choosing a statistical test

The dataset presented contains results from a number of participants in the Stroop test. Each row of the dataset contains the performance for one participant; with the first number their results on the congruent task and the second number their performance on the incongruent task.

Hence, the experiment has a within subject design as the same group of subjects receives different treatments. There are two conditions:

1. First condition/treatment: The subjects are made to perform the congruent task (In the congruent words condition, the words being displayed are color words whose names match the colors in which they are printed)
2. Second condition/treatment: The subjects are made to perform the incongruent task (In the incongruent words condition, the words displayed are color words whose names do not match the colors in which they are printed)

Hence, we have paired data sets that are dependent on each other.

Congruent data: 12.079, 16.791, 9.564...

Incongruent data: 19.278, 18.741, 21.214...

The two datasets are dependent on each other.

Note that: That data is sample data and hence, we cannot determine the population parameters.

Statistical Test: Considering the above factors, we have to employ dependent t-tests for paired samples

Appropriate set of Hypotheses:

1. Participants are slower to properly identify the font color when the font color used to produce colour names different from the font color. That is, participants are slower to identify red coloured font when it spells the word blue.
   On the other hand,

2. There might not be enough evidence to show that incongruent task takes more time than the congruent task.

Although from a scientific standpoint,

The "Speed of Processing" hypothesis suggests that word processing is much faster than color processing. Thus, in a situation of incongruence between words and colors, when the task is to report the color, the word information arrives at the decision process stage earlier than the color information and results in processing confusion. Hence, congruent data would be lesser than incongruent data.

Statistical representation of hypothesis:

The given dataset contains results from 24 participants.

Each row of the dataset contains the time in seconds for each participant, with the first number being the results on the congruent task and the second number being the result on the incongruent task.

We want to see if there is a significant difference between these two numbers:

Null hypothesis: There is no significant difference; i.e. they are almost equal.

Extending it to the entire population:

$$H_0: \quad \mu_c - \mu_I = 0$$

Alternate hypothesis: There is a significant difference between the population means of congruent and incongruent time.

$$H_A: \quad \mu_c - \mu_I \neq 0$$

Note:

μ_c = Population mean for the Congruent Data

μ_I = Population mean for the Incongruent Data

2.3 Descriptive statistics

a. Central tendency Measure

For Congruent Data,

$\bar{x}_c$ = Mean time for a sample of Congruent data

$\bar{x}_c$ = 14.051125

For Incongruent Data,

$\bar{x}_I$ = Mean time for a sample of Incongruent data

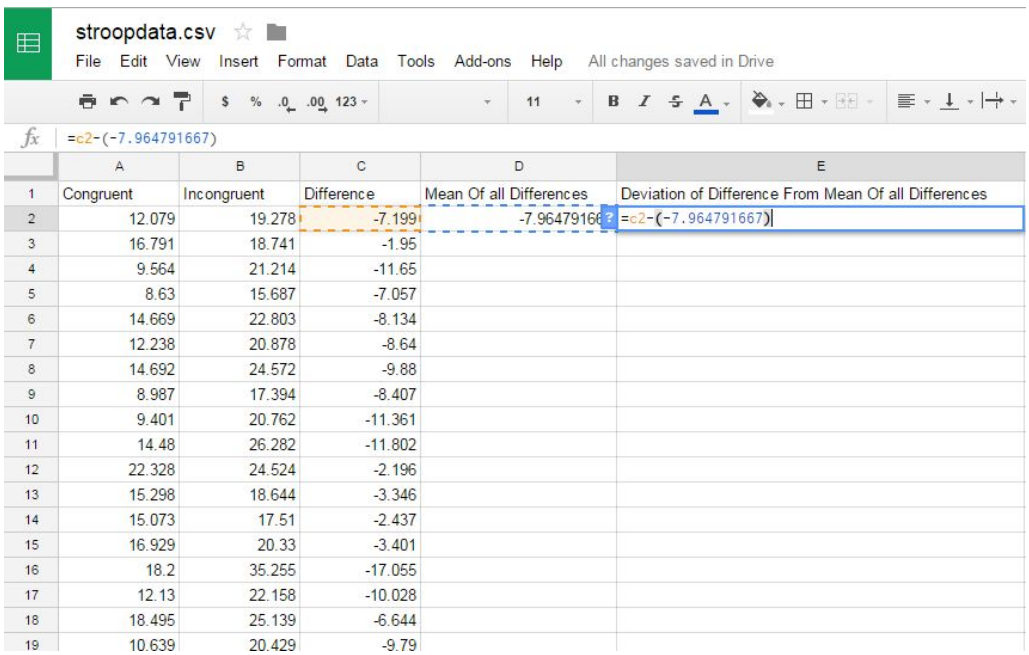
$\bar{x}_I$ = 22.01591667

b. Variability Measure

To find the standard deviation of the each difference:

Step 1:

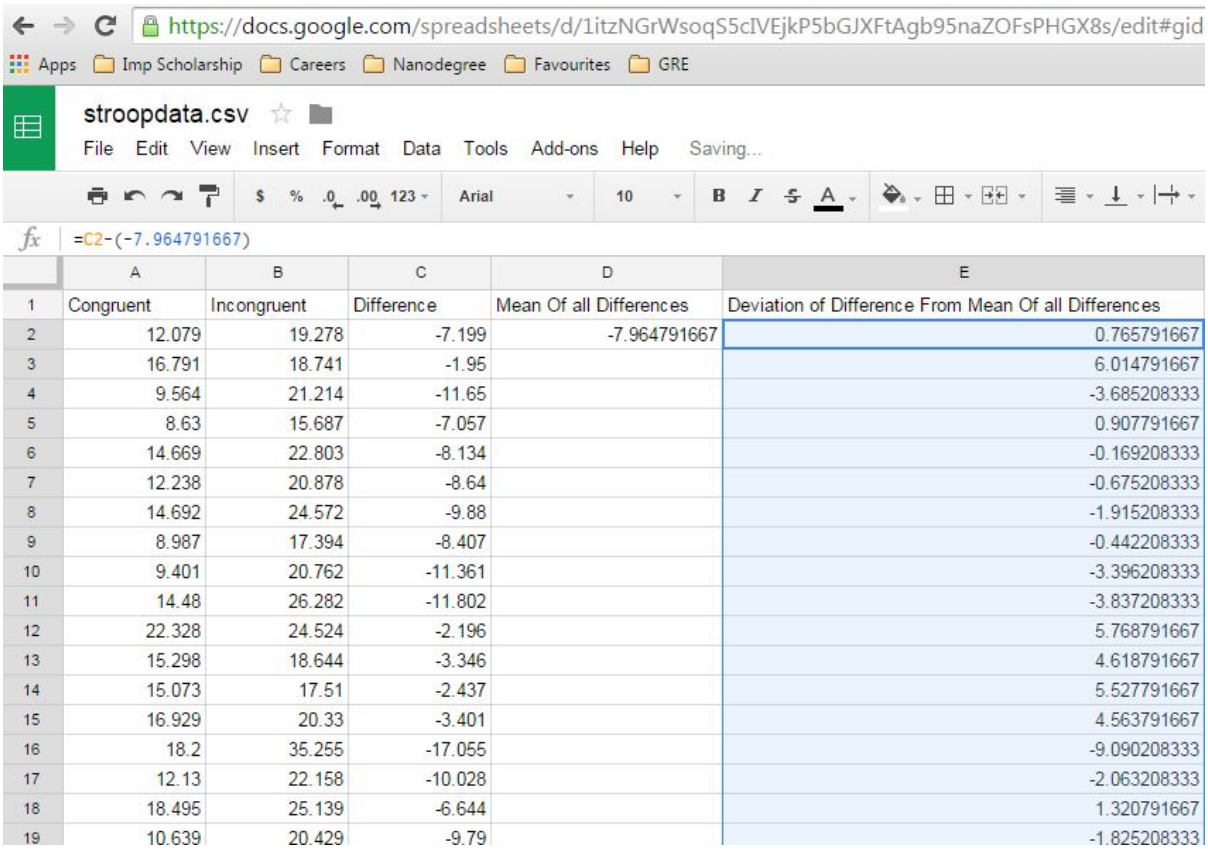
Find the difference between the incongruent and congruent data pairs; take out the mean of this difference



	A	B	C	D	E
1	Congruent	Incongruent	Difference	Mean Of all Differences	Deviation of Difference From Mean Of all Differences
2	12.079	19.278	-7.199	-7.964791667	=c2-(-7.964791667)
3	16.791	18.741	-1.95		
4	9.564	21.214	-11.65		
5	8.63	15.687	-7.057		
6	14.669	22.803	-8.134		
7	12.238	20.878	-8.64		
8	14.692	24.572	-9.88		
9	8.987	17.394	-8.407		
10	9.401	20.762	-11.361		
11	14.48	26.282	-11.802		
12	22.328	24.524	-2.196		
13	15.298	18.644	-3.346		
14	15.073	17.51	-2.437		
15	16.929	20.33	-3.401		
16	18.2	35.255	-17.055		
17	12.13	22.158	-10.028		
18	18.495	25.139	-6.644		
19	10.639	20.429	-9.79		

Step 2:

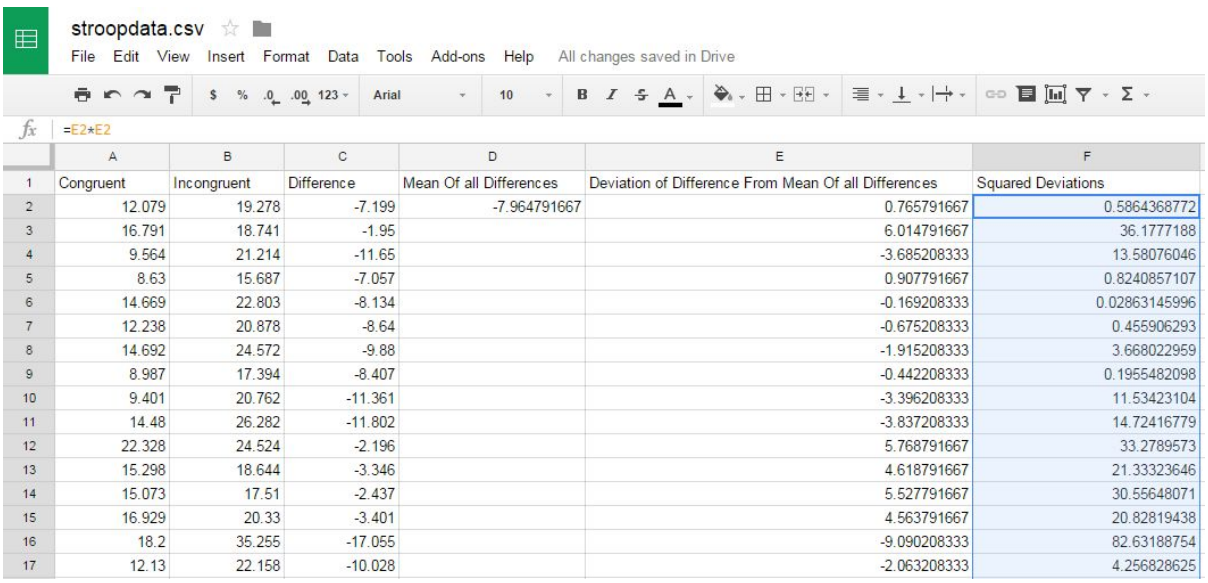
Find the deviation between the mean and each of the difference values



	A	B	C	D	E
1	Congruent	Incongruent	Difference	Mean Of all Differences	Deviation of Difference From Mean Of all Differences
2	12.079	19.278	-7.199	-7.964791667	0.765791667
3	16.791	18.741	-1.95		6.014791667
4	9.564	21.214	-11.65		-3.685208333
5	8.63	15.687	-7.057		0.907791667
6	14.669	22.803	-8.134		-0.169208333
7	12.238	20.878	-8.64		-0.675208333
8	14.692	24.572	-9.88		-1.915208333
9	8.987	17.394	-8.407		-0.442208333
10	9.401	20.762	-11.361		-3.396208333
11	14.48	26.282	-11.802		-3.837208333
12	22.328	24.524	-2.196		5.768791667
13	15.298	18.644	-3.346		4.618791667
14	15.073	17.51	-2.437		5.527791667
15	16.929	20.33	-3.401		4.563791667
16	18.2	35.255	-17.055		-9.090208333
17	12.13	22.158	-10.028		-2.063208333
18	18.495	25.139	-6.644		1.320791667
19	10.639	20.429	-9.79		-1.825208333

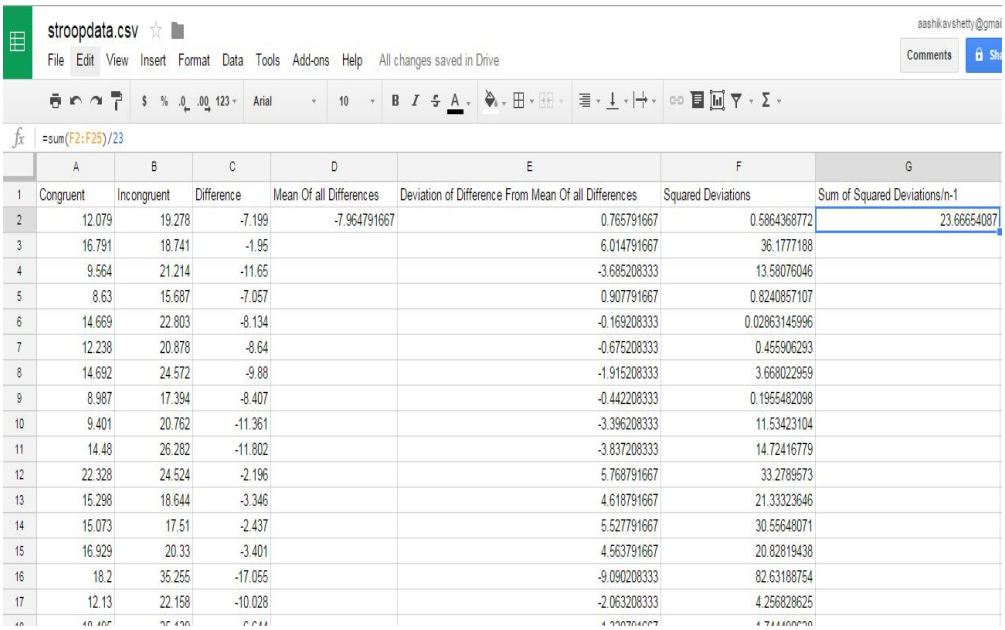
Step 3:

Find the squared deviation



	A	B	C	D	E	F
1	Congruent	Incongruent	Difference	Mean Of all Differences	Deviation of Difference From Mean Of all Differences	Squared Deviations
2	12.079	19.278	-7.199	-7.964791667	0.765791667	0.5864368772
3	16.791	18.741	-1.95		6.014791667	36.1777188
4	9.564	21.214	-11.65		-3.685208333	13.58076046
5	8.63	15.687	-7.057		0.907791667	0.8240857107
6	14.669	22.803	-8.134		-0.169208333	0.02863145996
7	12.238	20.878	-8.64		-0.675208333	0.455906293
8	14.692	24.572	-9.88		-1.915208333	3.668022959
9	8.987	17.394	-8.407		-0.442208333	0.1955482098
10	9.401	20.762	-11.361		-3.396208333	11.53423104
11	14.48	26.282	-11.802		-3.837208333	14.72416779
12	22.328	24.524	-2.196		5.768791667	33.2789573
13	15.298	18.644	-3.346		4.618791667	21.33323646
14	15.073	17.51	-2.437		5.527791667	30.55648071
15	16.929	20.33	-3.401		4.563791667	20.82819438
16	18.2	35.255	-17.055		-9.090208333	82.63188754
17	12.13	22.158	-10.028		-2.063208333	4.256828625

Step 4: Find the mean squared deviation by using the formula sum of the squared deviation divided by (n-1)



stroopdata.csv

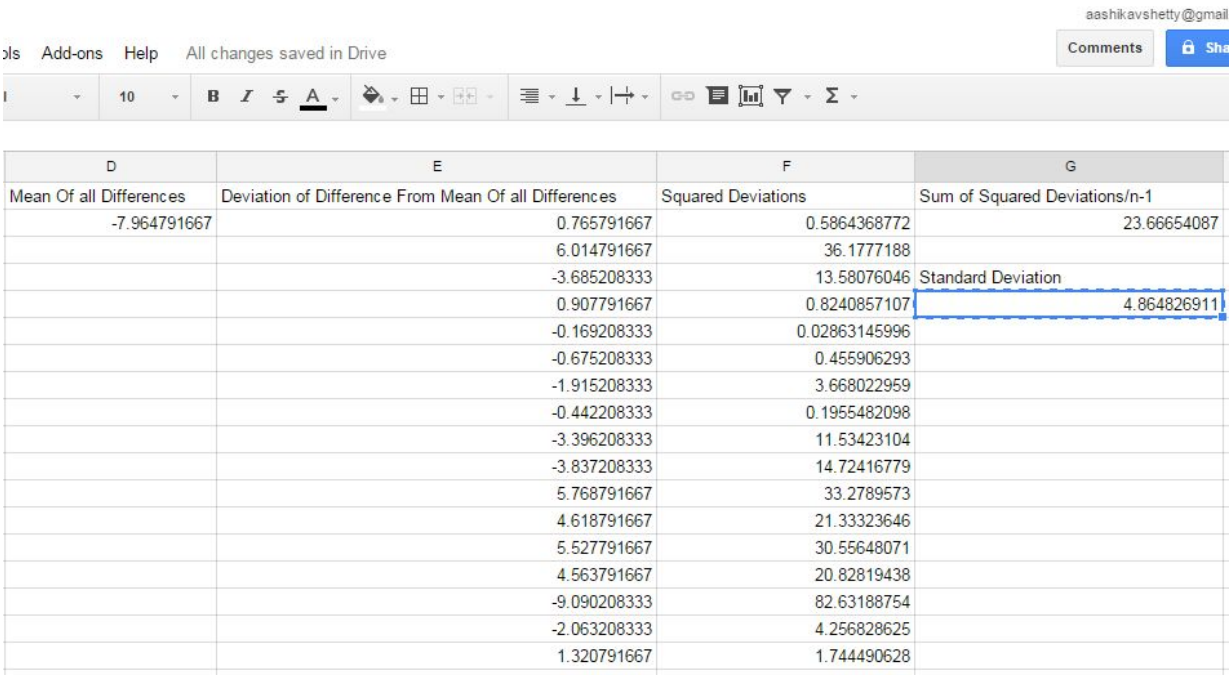
File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

=sum(F2:F25)/23

	A	B	C	D	E	F	G
1	Congruent	Incongruent	Difference	Mean Of all Differences	Deviation of Difference From Mean Of all Differences	Squared Deviations	Sum of Squared Deviations/n-1
2	12.079	19.278	-7.199	-7.964791667	0.765791667	0.5864368772	23.66654087
3	16.791	18.741	-1.95		6.014791667	36.1777188	
4	9.564	21.214	-11.65		-3.685208333	13.58076046	
5	8.63	15.687	-7.057		0.907791667	0.8240857107	
6	14.669	22.803	-8.134		-0.169208333	0.02863145996	
7	12.238	20.878	-8.64		-0.675208333	0.455906293	
8	14.692	24.572	-9.88		-1.915208333	3.668022959	
9	8.987	17.394	-8.407		-0.442208333	0.1955482098	
10	9.401	20.762	-11.361		-3.396208333	11.53423104	
11	14.48	26.282	-11.802		-3.837208333	14.72416779	
12	22.328	24.524	-2.196		5.768791667	33.2789573	
13	15.298	18.644	-3.346		4.618791667	21.33323646	
14	15.073	17.51	-2.437		5.527791667	30.55648071	
15	16.929	20.33	-3.401		4.563791667	20.82819438	
16	18.2	35.255	-17.055		-9.090208333	82.63188754	
17	12.13	22.158	-10.028		-2.063208333	4.256828625	

Step 5:

Find the square root of the mean squared deviation



aashikavshetty@gmail

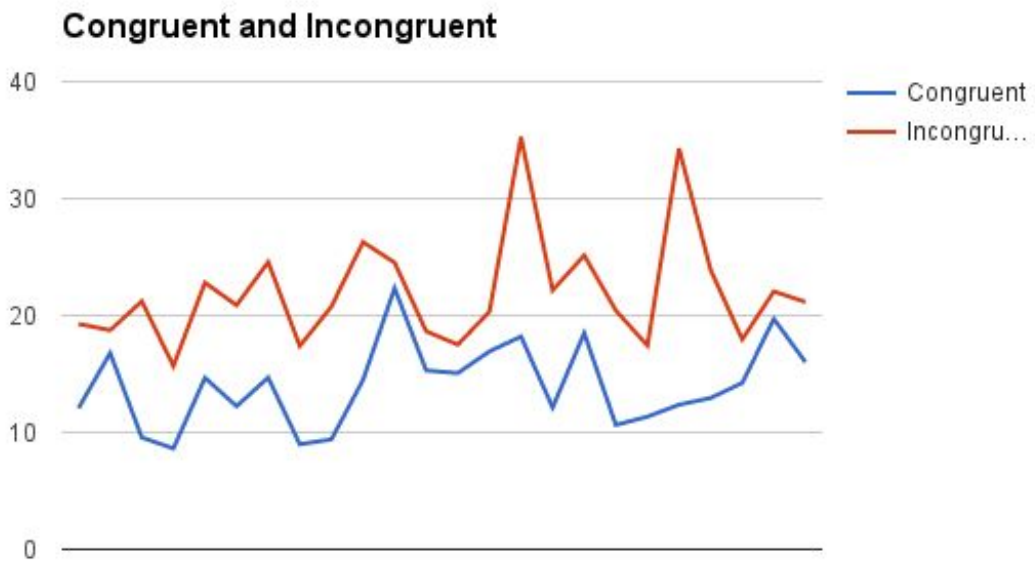
ols Add-ons Help All changes saved in Drive

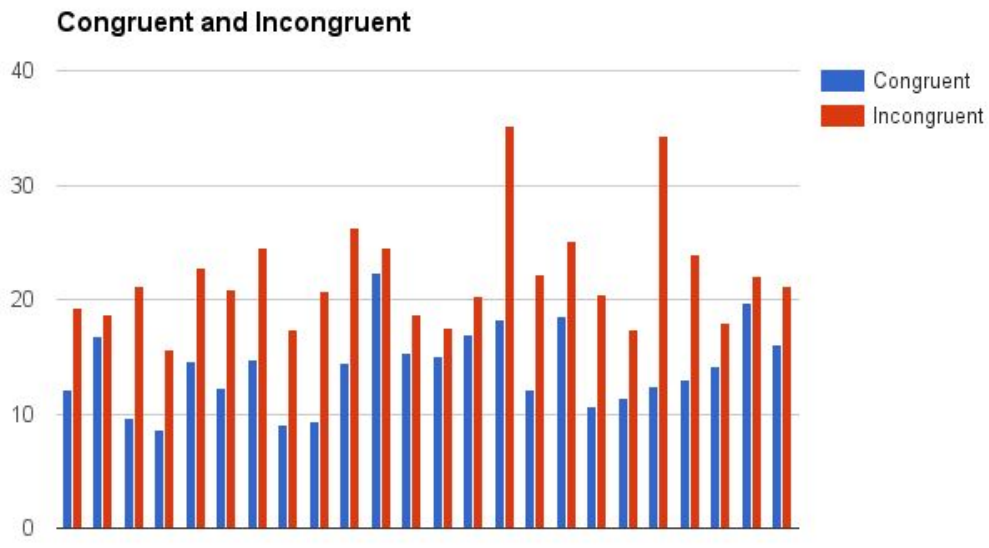
Comments

D	E	F	G
Mean Of all Differences	Deviation of Difference From Mean Of all Differences	Squared Deviations	Sum of Squared Deviations/n-1
-7.964791667	0.765791667	0.5864368772	23.66654087
	6.014791667	36.1777188	
	-3.685208333	13.58076046	Standard Deviation
	0.907791667	0.8240857107	4.864826911
	-0.169208333	0.02863145996	
	-0.675208333	0.455906293	
	-1.915208333	3.668022959	
	-0.442208333	0.1955482098	
	-3.396208333	11.53423104	
	-3.837208333	14.72416779	
	5.768791667	33.2789573	
	4.618791667	21.33323646	
	5.527791667	30.55648071	
	4.563791667	20.82819438	
	-9.090208333	82.63188754	
	-2.063208333	4.256828625	
	1.320791667	1.744490628	

Therefore, S= 4.86482691

2.4 Data Plot and the interpretation





From analysing the plots above, the congruent data value is always lesser than the corresponding to the incongruent data value.

2.5 Statistical test and analysis

We have to perform dependent t-tests for paired samples:

Known parameters:

Mean time for a sample of Congruent data:

$\overline{x}_c$ = 14.051125

Mean time for a sample of Incongruent data:

$\overline{x}_I$ = 22.01591667

S= 4.864826911

Statistics:

Point Estimate = $\overline{x}_c$ - $\overline{x}_I$ = 14.051125 -22.01591667 = -7.96479167

By the formula for t we have,

$t = (\mu_c - \mu_I)/(S/\sqrt{n})$

$t = -7.96479167/(4.864826911/\sqrt{24})$

t = -7.96479167/0.9930286349

t =-8.02070694648

$t_{statistics}$ = -8.02070694648

For an α *level* of 0.05, the value of $t_{critical}$,

$t_{critical}$= -2.069 or 2.069

Since, the $t_{statistics}$ is way past the value of $t_{critical}$ in the negative region , the null hypothesis is rejected and μi is significantly lesser than μc. This means that the participants took significantly lesser time for the congruent task as compared to the incongruent task.

Cohen's d: (For the effect size measure)

Cohen's d = mean difference/ standard deviation

= -7.96479167/ 4.864826911

Cohen's d= -1.63721994959

Confidence Interval for the mean population deviation:

$CI = M_D \pm t_{critical}(S/\sqrt{n})$

Confidence Interval =-6.0016748991 to -9.9279084409

This tells us that on an average the participants will take 6.0016748991 to 9.9279084409 seconds more when presented with an incongruent data set as opposed to when presented with a congruent data set.

Word processing by human brain, logically speaking is much faster than color processing. Thus, in a situation of incongruence between words and colors, when the task is to report the color, the word information arrives at the decision process stage earlier than the color information and results in processing confusion. Hence,we expect the congruent task would require lesser time than incongruent task.

The actual results do match up with our expectations.

2.6 Extension of the experiment

Reason for this observation:

There is a lag in the brain's ability to recognize the color of the word since the brain reads words faster than it recognizes colors. Word processing often happens much faster than color processing.

Also, reading out the color is not reflexive or automatic i.e. it takes a lot of attention, and concentration.

Alternative or similar task that would result in a similar effect:

When the time for reading names of colors had been compared with the time for naming colors themselves, we observed an interference effect.

This interference i.e. the increase in time for reacting to words caused by the presence of conflicting color stimuli is taken as the measure of the interference of color stimuli upon reading words. The increase in the time for reacting to colors caused by the presence of conflicting word stimuli is taken as the measure of the interference of word stimuli upon naming colors.

The effect of the interference is likely to reduce after practice.

Similar Experiments:

1. The subjects associated each of a series of numbers with striking a particular key on the typewriter with a particular finger; then the keys were changed so that four of the numbers had to be written with fingers other than those formerly used to write them. In the other experiment the subjects were trained to react with the right hand to 'red' and with the left hand to 'blue.' Then the stimuli were interchanged. In the former experiment an interference was found which decreased rapidly with practice. In the latter experiment the interference was overbalanced by the practice effect.

(Taken from **STUDIES OF INTERFERENCE IN SERIAL VERBAL REACTIONS**

J. Ridley Stroop[1] (1935)

George Peabody College

First published in *Journal of Experimental Psychology*, *18*, 643-662.)

3. References:

https://en.wikipedia.org/wiki/Stroop_effect

https://faculty.washington.edu/chudler/java/ready.html

http://www.scribd.com/doc/111646449/Stroop-Effect#scribd

http://www.slideshare.net/mkw1025kenneth/stroop-effect

http://www.rit.edu/cla/gssp400/sbackground.html

https://docs.google.com/document/d/1-OkpZLjG_kX9J6LIQ5IltsqMzVWjh36QpnP2RYpVdPU/pub?embedded=True

http://psychclassics.yorku.ca/Stroop/