UDACITY PROJECT

SUBMITTED BY: Akanksha Goel

DATE: 16 May,2017

1. **What is our independent variable? What is our dependent variable?**
   Dependent Variable: The time it takes to name the ink colors in equally-sized lists.
   Independent Variable : The two independent conditions that is chosen whether the word and color is congruent or incongruent.
   Source: Given.
2. **What is an appropriate set of hypotheses for this task? What kind of statistical test do you expect to perform? Justify your choices.**

   Ho (null hypothesis) : that there will be no affect in the time taken while reading incongruent words .

   μ(congruent) $\geq \mu$(incongruent)

   (population mean of time taken will remain same to population mean when used congruent or incongruent set.)

   Ha (alternative hypothesis): that there will be increase in time taken while reading incongruent words i.e

   μ(congruent) $< \mu$(incongruent)

   (population mean will increase as you use the incongruent set instead of congruent set)

   In this **t test is used** to analyse hypothesis as we **don't have population standard deviation σ** and don't have **sample size greater than 30**.

   We assume that the distributions are Gaussian.

   This test is applied on these **dependent samples** because same population is used to read congruent set and incongruent set. **Time is measured in two conditions on the same population .**

   This hypothesis testing will be conducted using **one-tail test**.

   This statistical analysis is being done because time used in reading incongruent words is greater then time required to read congruent words as per the stroop's effect.

   As **correlation cannot be considered as causation of effect** therefore inferential statistics is being used to infer results using experiment.

   Source: https://en.wikipedia.org/wiki/Stroop_effect
3. **Report some descriptive statistics regarding this dataset. Include at least one measure of central tendency and at least one measure of variability.**

   Measure of central tendency : Mean(congruent) = Average time taken to read congruent words = $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$.

Mean(incongruent) = Average time taken to read incongruent

$$\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i.$$

words =

Measure of variability :

The **sample standard deviation formula** is:

$$s = \sqrt{\frac{\sum(X - \bar{X})^2}{n-1}}$$

where,

s = sample standard deviation

$\sum$ = sum of...

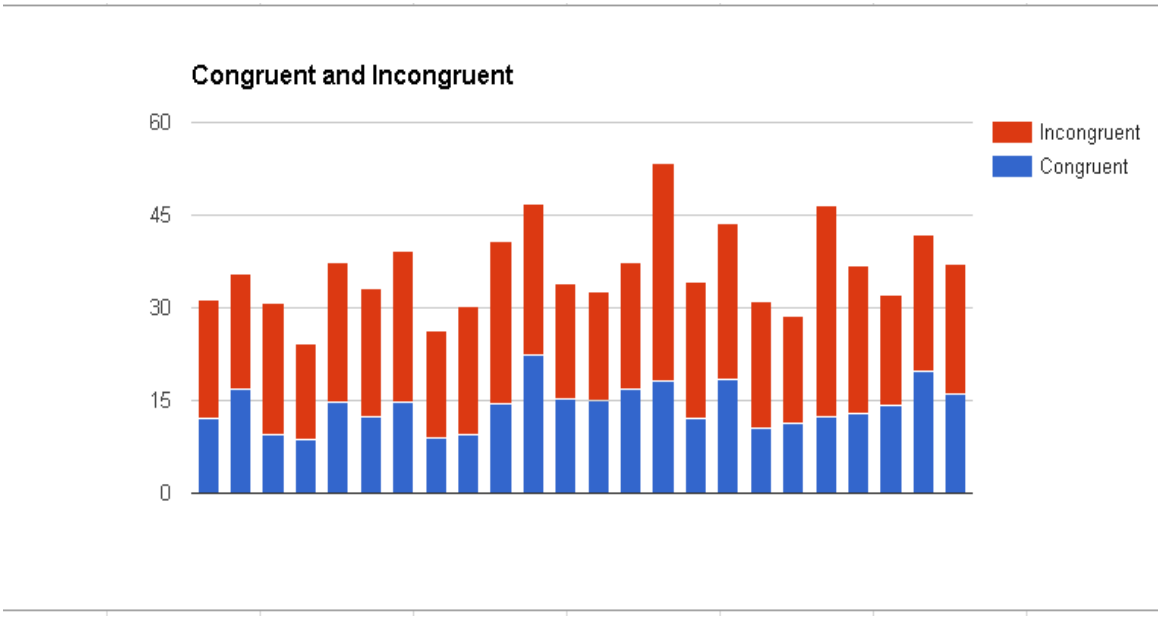
$\bar{X}$ = sample mean

n = number of scores in sample.

	A	B	C	D	E	F	G	H
1	Congruent(c)	Incongruent(i)	deviation of (C) (x-xbar)	deviation of (I) (x-xbar)	squared deviation (C) (x-xbar)^2	squared deviation (I) (x-xbar)^2		
2	12.079	19.278	-1.972125	-2.737916667	3.889277016	7.496187674	Mean(c)-->	14.051125
3	16.791	18.741	2.739875	-3.274916667	7.506915016	10.72507917	Mean(i)-->	22.01591667
4	9.564	21.214	-4.487125	-0.801916667	20.13429077	0.64307034	variance(c)-->	12.66902907
5	8.63	15.687	-5.421125	-6.328916667	29.38859627	40.05518617	variance(i)-->	23.01175704
6	14.669	22.803	0.617875	0.787083333	0.381769516	0.619500174	standard dev(c)--	3.559357958
7	12.238	20.878	-1.813125	-1.137916667	3.287422266	1.29485434	standard dev(i)-->	4.797057122
8	14.692	24.572	0.640875	2.556083333	0.410720766	6.533562007		
9	8.987	17.394	-5.064125	-4.621916667	25.64536202	21.36211367		
10	9.401	20.762	-4.650125	-1.253916667	21.62366252	1.572307007		
11	14.48	26.282	0.428875	4.266083333	0.183933766	18.19946701		
12	22.328	24.524	8.276875	2.508083333	68.50665977	6.290482007		
13	15.298	18.644	1.246875	-3.371916667	1.554697266	11.36982201		
14	15.073	17.51	1.021875	-4.505916667	1.044228516	20.30328501		
15	16.929	20.33	2.877875	-1.685916667	8.282164516	2.842315007		
16	18.2	35.255	4.148875	13.23908333	17.21316377	175.2733275		
17	12.13	22.158	-1.921125	0.142083333	3.690721266	0.020187674		
18	18.495	25.139	4.443875	3.123083333	19.74802502	9.753649507		
19	10.639	20.429	-3.412125	-1.586916667	11.64259702	2.518304507		
20	11.344	17.425	-2.707125	-4.590916667	7.328525766	21.07651584		
21	12.369	34.288	-1.682125	12.27208333	2.829544516	150.6040293		
22	12.944	23.894	-1.107125	1.878083333	1.225725766	3.527197007		
23	14.233	17.96	0.181875	-4.055916667	0.033078516	16.45046001		
24	19.71	22.058	5.658875	0.042083333	32.02286627	0.001771007		
25	16.004	21.157	1.952875	-0.858916667	3.813720766	0.73773784		

.

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.



X Axis represent users.

Y Axis represents time taken by respective user.

For sample:

Time taken in reading incongruent words > Time taken while reading congruent words.
For population:
The chart shows that time taken while reading incongruent words may be more than that of reading congruent words for user.

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?**

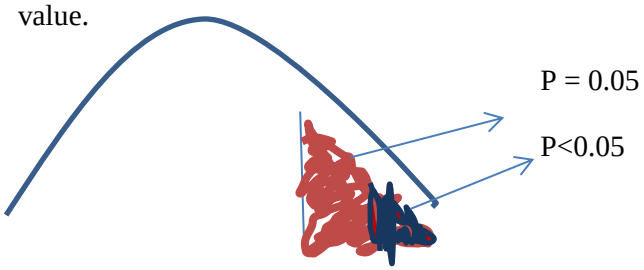
- Using t-distribution we check that the independent variable will have an effect on dependent variable using 1 tail test.
- Let the α value for distribution be 0.05.
- n=sample size=24
- **degree of freedom**: 23
- Finding the **t-critical value**:1.714 (using t-table)
- Mean Difference : 7.964
- Standard Deviation of difference :4.86

J5 =J4/SQRT(24)

	A	B	C	D	E	I	J
1	Congruent(c)	Incongruent(i)	Difference(i-c)	deviation of (x-xbar)	squared deviation (I) (x-xbar)^2		
2	12.079	19.278	7.199	-0.765791667	0.586436877	Mean difference-->	7.964792
3	16.791	18.741	1.95	-6.014791667	36.17771879	variance -->	23.66654
4	9.564	21.214	11.65	3.685208333	13.58076046	Standard Deviation-->	4.864827
5	8.63	15.687	7.057	-0.907791667	0.82408571	Standard Error-->	0.993029
6	14.669	22.803	8.134	0.169208333	0.02863146	t-value-->	8.020707
7	12.238	20.878	8.64	0.675208333	0.455906293		
8	14.692	24.572	9.88	1.915208333	3.66802296		
9	8.987	17.394	8.407	0.442208333	0.19554821		
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	33.27895729		
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	4.256828627		
18	18.495	25.139	6.644	-1.320791667	1.744490627		
19	10.639	20.429	9.79	1.825208333	3.33138546		
20	11.344	17.425	6.081	-1.883791667	3.548671043		
21	12.369	34.288	21.919	13.95420833	194.7199302		
22	12.944	23.894	10.95	2.985208333	8.911468793		
23	14.233	17.96	3.727	-4.237791667	17.95887821		
24	19.71	22.058	2.348	-5.616791667	31.54834863		
25	16.004	21.157	5.153	-2.811791667	7.906172377		

stroopdata

- standard Error = $\frac{Standard\ Deviation}{\sqrt{n}} = \frac{4.864}{\sqrt{24}} = 0.9930$
- t statistical value = $\frac{mean(incongruent) - mean(incongruent)}{standard\ error} = \frac{7.964}{.99} = 8.020$
- P-value for this t value must be less then .05 because t value is greater then t-critical value.



1.714 8.02

- As t (statistical value) is greater than t-critical value therefore
  The t value lies in critical region therefore we will reject null hypothesis and accept alternative Hypothesis .
  μ(congruent) <μ(incongruent) (population mean)
  Hence it is evident that bringing incongruent words increase the time taken by user.
  Yes, the result came out to be totally different to what we expected. *The time taken while reading incongruent words is more than the time required to read congruent words .*
- For finding confidence interval let α be 0.05.
  Now t-critical value be ± 2.069 .(using t-table)
  Standard error : .9930
  Mean difference : 22.01 – 14.05 = 7.96 .
  95% confidence interval : mean difference ± (t-critical * Standard Error) = (5.91, 10.008).

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!**

There are two theories that may explain the Stroop effect:

1. Speed of Processing Theory: the interference occurs because words are read faster than colors are named.
2. Selective Attention Theory: the interference occurs because naming colors requires more attention than reading words.

Alternative Task : Colouring the words half with one colour and other half with different colour and reading the word instead of colour