

## Test a Perceptual phenomenon (Stroop Data)

### Question 1: Identify variables in the experiment.

The independent variable is the congruency of the colour and the text.

The dependent variable is the Duration (time) taken to recognise the colours of the text.

### Q2a: Null and alternative hypotheses are clearly stated in words and mathematically. Symbols in the mathematical statement are defined.

Null hypothesis ( $H_0$ ): The mean time taken by the participants to identify the font colour under both the conditions (*congruent data* and *incongruent data*) are not different.

$$H_0: \mu_x = \mu_y$$

Alternative hypothesis ( $H_A$ ): The mean time taken by the participants to identify the font colour under both the conditions (*congruent data* and *incongruent data*) are different.

$$H_A: \mu_x \neq \mu_y$$

Here 'x' is used to refer to the sample of *congruent data* whereas as 'y' is used to refer to *incongruent data*.

Thus the Null hypothesis states that there is no difference in the mean time taken by participants to say out loud the colours of the font in *congruent data* as compared to the *incongruent data*.

The Alternative hypothesis states that there is a difference in the mean time taken by participants to say out loud the colours of the font in the *congruent data* as compared to the *incongruent data*.

### Q2b: A statistical test is proposed which will distinguish the proposed hypotheses. Any assumptions made by the statistical test are addressed.

The statistical test used is dependent samples t-test with two-tailed distribution since we want to find out whether there is a change in the time taken by participants to read out the colours in the *congruent data* as compared to the *incongruent data*.

The dataset given have been considered as samples since the data size is less than 25 and I have used 95% confidence level for conducting the t-test.

The distributions of the samples are assumed to be normal.

**Q3: Descriptive statistics, including at least one measure of centrality and one measure of variability, have been computed for the dataset's groups.**

Below I have calculated the measure of central tendency (mean) for both congruent and incongruent data. I have also calculated a measure of variability (standard deviation) for the given data.

<b>Congruent (x)</b>	<b>Incongruent (y)</b>	<b>Difference= x - Mean(x)</b>	<b>Difference= y - Mean(y)</b>	<b>Squared difference (x)</b>	<b>Squared Difference (y)</b>
12.079	19.278	-1.972125	-2.737916667	3.889277016	7.496187674
16.791	18.741	2.739875	-3.274916667	7.506915016	10.72507917
9.564	21.214	-4.487125	-0.801916667	20.13429077	0.64307034
8.63	15.687	-5.421125	-6.328916667	29.38859627	40.05518617
14.669	22.803	0.617875	0.787083333	0.381769516	0.619500174
12.238	20.878	-1.813125	-1.137916667	3.287422266	1.29485434
14.692	24.572	0.640875	2.556083333	0.410720766	6.533562007
8.987	17.394	-5.064125	-4.621916667	25.64536202	21.36211367
9.401	20.762	-4.650125	-1.253916667	21.62366252	1.572307007
14.48	26.282	0.428875	4.266083333	0.183933766	18.19946701
22.328	24.524	8.276875	2.508083333	68.50665977	6.290482007
15.298	18.644	1.246875	-3.371916667	1.554697266	11.36982201
15.073	17.51	1.021875	-4.505916667	1.044228516	20.30328501
16.929	20.33	2.877875	-1.685916667	8.282164516	2.842315007
18.2	35.255	4.148875	13.23908333	17.21316377	175.2733275
12.13	22.158	-1.921125	0.142083333	3.690721266	0.020187674
18.495	25.139	4.443875	3.123083333	19.74802502	9.753649507
10.639	20.429	-3.412125	-1.586916667	11.64259702	2.518304507
11.344	17.425	-2.707125	-4.590916667	7.328525766	21.07651584
12.369	34.288	-1.682125	12.27208333	2.829544516	150.6040293
12.944	23.894	-1.107125	1.878083333	1.225725766	3.527197007
14.233	17.96	0.181875	-4.055916667	0.033078516	16.45046001
19.71	22.058	5.658875	0.042083333	32.02286627	0.001771007
16.004	21.157	1.952875	-0.858916667	3.813720766	0.73773784

#### **Measure of Variability:**

Standard Deviation (x) =  $\sqrt{\text{Sum of Squared Differences}(x)/(n-1)}$  = 3.55935796

Standard Deviation (y) =  $\sqrt{\text{Sum of Squared Differences}(y)/(n-1)}$  = 4.79705712

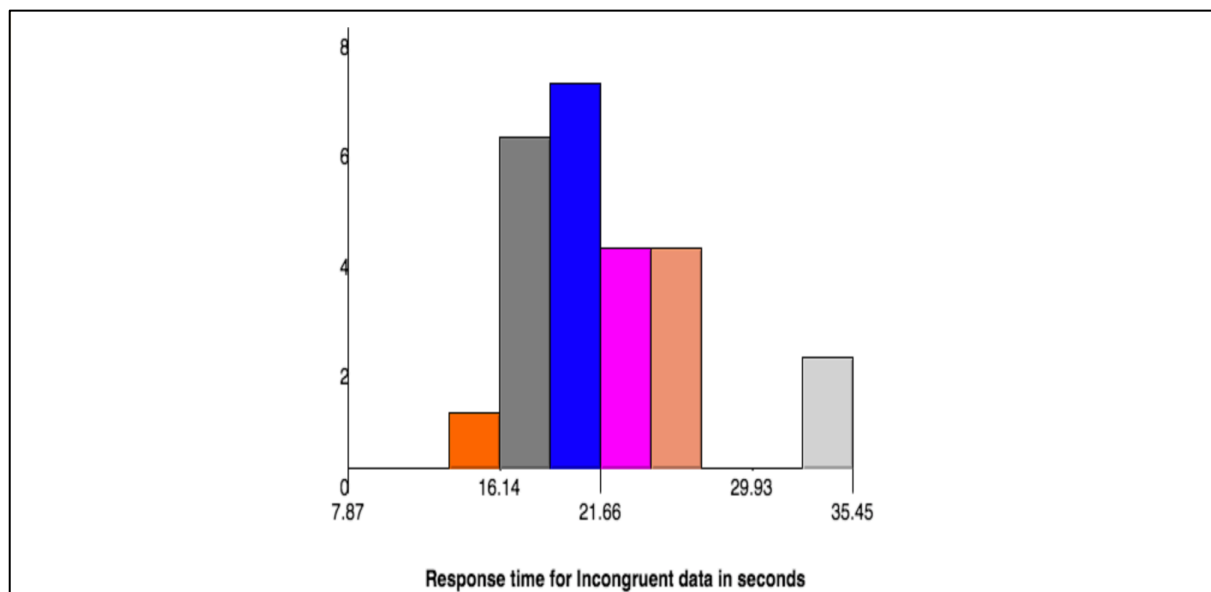
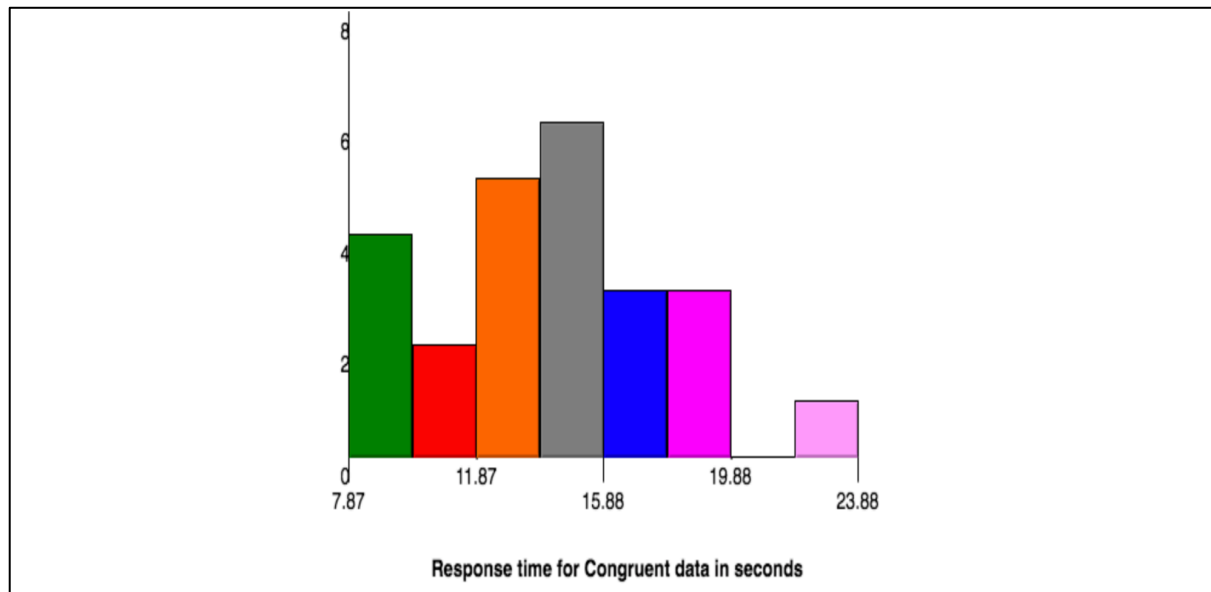
#### **Measure of Central tendency:**

n = 24

Mean (x) =  $(x_1+x_2+x_3+\dots+x_{24})/n$  = 14.051125

Mean (y) =  $(y_1+y_2+y_3+\dots+y_{24})/n$  = 22.0159167

**Q4:** One or two visualizations have been created that show off the data, including comments on what can be observed in the plot or plots.



In the congruent data histogram, the maximum number of observations lie between 11.87 seconds and 15.88 seconds.

In the incongruent data histogram the, the maximum number of observations lie between 16.14 seconds and 21.66 seconds.

Both the data are more or less normally distributed.

**Question 5: Perform the statistical test and interpret your results.**

Congruent (x)	Incongruent (y)	Difference (x-y)	Square of difference
12.079	19.278	-7.199	51.825601
16.791	18.741	-1.95	3.8025
9.564	21.214	-11.65	135.7225
8.63	15.687	-7.057	49.801249
14.669	22.803	-8.134	66.161956
12.238	20.878	-8.64	74.6496
14.692	24.572	-9.88	97.6144
8.987	17.394	-8.407	70.677649
9.401	20.762	-11.361	129.072321
14.48	26.282	-11.802	139.287204
22.328	24.524	-2.196	4.822416
15.298	18.644	-3.346	11.195716
15.073	17.51	-2.437	5.938969
16.929	20.33	-3.401	11.566801
18.2	35.255	-17.055	290.873025
12.13	22.158	-10.028	100.560784
18.495	25.139	-6.644	44.142736
10.639	20.429	-9.79	95.8441
11.344	17.425	-6.081	36.978561
12.369	34.288	-21.919	480.442561
12.944	23.894	-10.95	119.9025
14.233	17.96	-3.727	13.890529
19.71	22.058	-2.348	5.513104
16.004	21.157	-5.153	26.553409

Degrees of Freedom =  $n-1 = 24-1 = 23$

Mean (x) = 14.051125

Mean (y) = 22.0159167

Point estimate = Mean (x) - Mean (y) = -7.9647917

Standard Deviation of Differences (s) =  $\sqrt{\sum(x-\text{Mean}(x))^2/(n-1)} = 4.86482691$

Mean Difference = -7.9647917

t-statistic = Mean difference /  $(s/\sqrt{n}) = -8.0207069$

t-critical value at 95% confidence level for 2 tailed distribution =  $\pm 2.069$

**My observations:**

- The t-statistic value (-8.021) is less than the t-critical value ( $\pm 2.069$ ) for a two tail test with confidence level of 95%. The t-statistic lies in the critical region. Therefore we can reject the null hypothesis.
- Therefore we observe that the average time taken by the participants to identify and speak out loud the font colour for incongruent data is greater than the average time taken to identify the font colour in congruent data.

**Additional resources used:**

- Histogram data in the excel file 'stroop data calculation' is constructed from the website <http://www.shodor.org/interactivate/activities/Histogram/>
- T-Table used from the one given by Udacity in Lesson 26 of Descriptive Statistics. <https://s3.amazonaws.com/udacity-hosted-downloads/t-table.jpg>