10th June 2016

DATA ANALYST NANODEGREE

PROJECT P1 : TEST A PERPETUAL PHENOMENON

Background Information

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.

Questions For Investigation with answers and statistical proof

1. What is our independent variable? What is our dependent variable?

   The **independent variable** in this experiment is whether the word and color are the same or different that is if they are congruent or incongruent.

The **dependent variable** is the reaction time taken to name the colour in each condition.

2. What is an appropriate set of hypotheses for this task? What kind of statistical test do you expect to perform? Justify your choices.

One of the hypothesis can be the difference in the average time taken to recognise the colors under congruent words condition and incongruent words condition, namely, the Stroop Effect is in existence.

Also, If we compare the means then we can infer that there is some difference between the two datasets color recognition times.

H_0: $\mu = 0$ (The difference between group population means is zero)

H_A: $\mu \neq 0$ (The difference between group population means is not zero)

A two tailed test will be used to decide whether we accept or reject the null hypothesis.
The dependent variable will be a normal distributed curve in both the conditions.

We can use a two-sided paired student t-test to verify for the following reasons:

1. We need to address the uncertainty in sample standard error resulted from the population standard deviation which is unknown
2. We are comparing the means of two dependent groups and the distributions are gaussian.
3. Both the conditions fall under the same topic with a sample size less than 30.

3. Report some descriptive statistics regarding this dataset. Include at least one measure of central tendency and at least one measure of variability.

Bessel’s correction has been used here .

Congruent dataset (Measure of central tendency):

- Mean : 14.051125
- Median : 14.3565

Incongruent dataset (Measure of central tendency) :

- Mean : 22.01591667
- Median : 21.0175

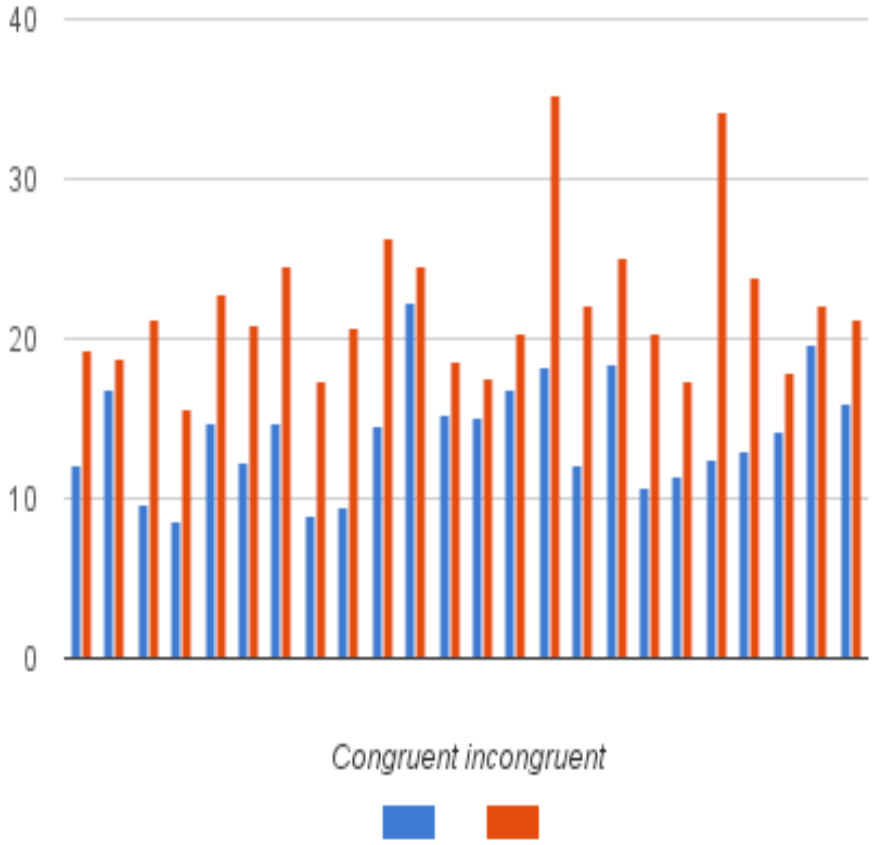
Congruent dataset (Measure of variability) :

- Standard Deviation : 3.559357958

Incongruent dataset (Measure of variability) :

- Standard Deviation : 4.797057122

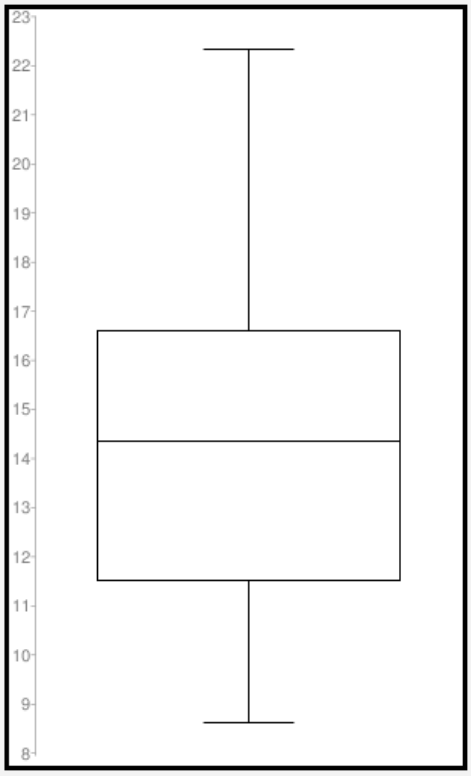
4. Provide one or two visualizations that show the distribution of the sample data. Write one or two sentences noting what you observe about the plot or plots.



Box plot for the congruent data set –

Sample size: 24
Median: 14.3565
Minimum: 8.63
Maximum: 22.328
First quartile: 11.52775
Third quartile: 16.59425
Interquartile Range: 5.0665
Outliers: none

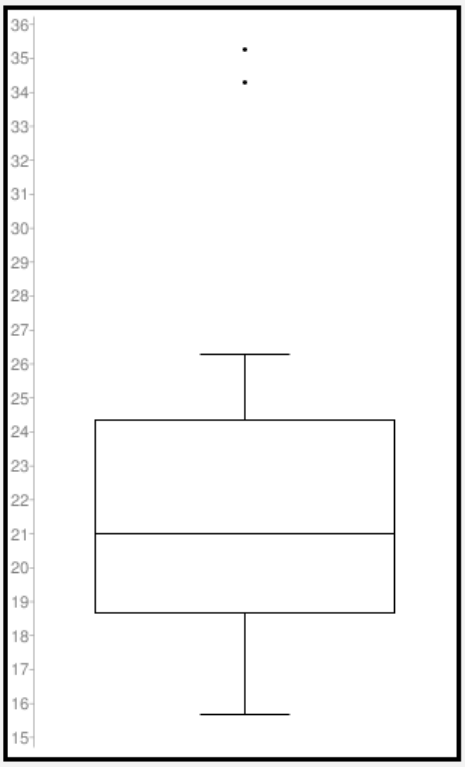
Mean
Median
Mode
Standard Deviation
Variance
Mean absolute deviation
Range
Interquartile range
Quartiles
All dispersion data



Box plot for the incongruent data set –

Sample size: 24
Median: 21.0175
Minimum: 15.687
Maximum: 35.255
First quartile: 18.66825
Third quartile: 24.3665
Interquartile Range: 5.69825
Outliers: 35.255 34.288

Mean
Median
Mode
Standard Deviation
Variance
Mean absolute deviation
Range
Interquartile range
Quartiles
All dispersion data



From, the above box plots we can say that :

1. The median of the incongruent data is higher than the median of the congruent data.
2. The congruent data has no outliers whereas the incongruent has 2 outliers.

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?

Mean of the difference of the incongruent and the congruent words conditions: **7.964791667**

The standard deviation (s) found from the difference between the incongruent and the congruent dataset: **4.86482691**

Standard Error : s / √n

s : standard deviation
n : size of sample (24)

Standard Error = 4.86482691 / √24 = **0.9930286347**

t-statistic = Mean Difference / Standard error
= **8.0207**

Consider, **α** (alpha confidence level) **= 0.05**

Since it's a **two tailed test** each side will have **2.5 % or 0.025** alpha level.

Degree of freedom = n-1, where n = 24
Therefore, **the degree of freedom = 23**

t-critical value = ± 2.069 dereived from the Students' t table.

The two-tailed P value equals **0.0505**
By conventional criteria, this difference is considered to be not quite statistically significant.

We **reject the null hypothesis** since the t-statistical value is greater than the t-critical value. Hence it falls in the critical region.

My expectation result was that the time taken to read incongruent words is more than the time taken to read congruent words.

The statistical result matches the expected result as we reject the null hypothesis it is safe to say that that the time taken to read incongruent words and the time taken to read congruent words is not the same.

References:

1. To design the box plot : http://www.alcula.com/calculators/statistics/box-plot/

2. The stoop test data set : https://docs.google.com/spreadsheets/d/1wSXlHAu7UValX-0Y6KX8okSt9XYBx6G1pIGbIhN3imQ/edit?usp=sharing

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