

Statistics: The Science of Decisions

Project Instructions

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.

Note:

The sample data is provided by Udacity and can be downloaded from site:

https://d17h27t6h515a5.cloudfront.net/topher/2016/September/57ce3363_stroopdata/stroopdata.csv

Analysis

Question 1: Identify variables in the experiment

The independent variable is the "the color of the ink in which the word is printed". There are two types of conditions: a congruent words condition, and an incongruent words condition.

The dependent variable is the time it takes to name the ink colors in equally-sized lists.

Question 2a: Establish hypotheses

Here we make a hypothesis that the incongruent test will increase the spoke out time than the congruent test significantly. And we use the 95% confidence level or Alpha level 5% to check the final result is significant or not. And we assume the incongruent test has more time than congruent, so that is one tail test at positive side of distribution and we hope to check if the result will locate in upper 5% region or not.

- The null hypothesis is: The congruent or incongruent will not affect spoke out time.
- And the reject hypothesis is: The incongruent will increase spoke out time.

Here we use $\mu_{\text{congruent}}$ for the population mean of congruent test; and use $\mu_{\text{incongruent}}$ for the population mean of incongruent.

Symbolically, these hypotheses would be expressed as:

$$H_0: \mu_{\text{congruent}} = \mu_{\text{incongruent}}$$

$$H_a: \mu_{\text{incongruent}} > \mu_{\text{congruent}}$$

Question 2b: Establish a statistical test

The Stroop test is a dependent test. To pick up some person to do test for congruent and incongruent. The sequence of congruent and incongruent is random. To record the time for congruent and incongruent test. Then calculate and compare sample data for congruent and incongruent.

Note:

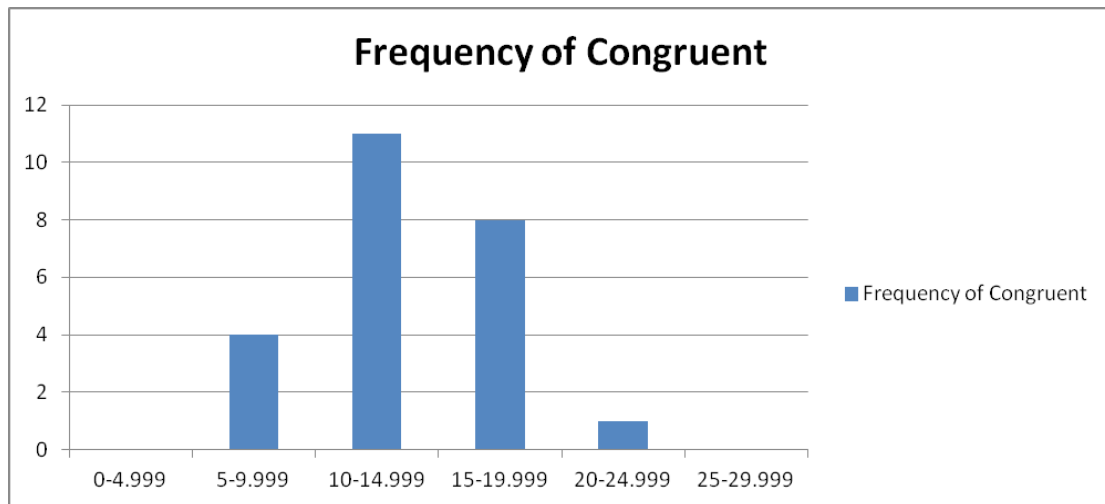
We use sample data for analysis. The sample data is downloaded from https://d17h27t6h515a5.cloudfront.net/topher/2016/September/57ce3363_stroopdata/stroopdata.csv

Question 3: Report descriptive statistics &

Question 4: Plot the data

Congruent data frequency is as below table, the unit is 5 second:

Time	Frequency of Congruent
0-4.999	0
5-9.999	4
10-14.999	11
15-19.999	8
20-24.999	1
25-29.999	0



The mode of frequency value is 11 which is in range 10-14.999;

The mean of sample is 14.05;

The median value is $(14.233+14.48)/2 = 14.36$;

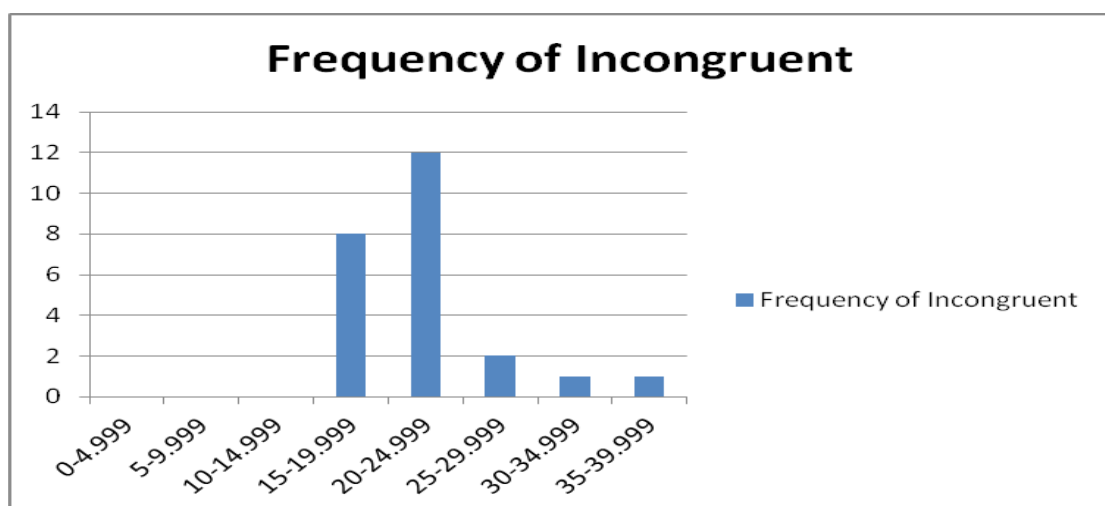
$IQR = Q3-Q1 = 16.004-12.079=3.925$;

And Mean in IQR is 13.85;

The standard deviation of samples is 3.56;

The Incongruent data frequency is as below table, unit is 5 second:

Time	Frequency of Incongruent
0-4. 999	0
5-9. 999	0
10-14. 999	0
15-19. 999	8
20-24. 999	12
25-29. 999	2
30-34. 999	1
35-39. 999	1



The mode of frequency value is 12 which is in range 20-24.999;

The mean of sample is 22.016;

The median value is $(20.878+21.157)/2=21.02$;

IQR=Q3-Q1= 23.894-18.741= 5.153;
 And the mean in IQR is 21.14;
 The standard deviation of samples is 4.797

The 95% confidence intervals of congruent sample is:
 $[\text{Mean}-2\sigma, \text{Mean}+2\sigma] = [14.05-2*3.56, 14.05+2*3.56] = [6.93, 21.14]$

The 95% confidence intervals of incongruent sample is:
 $[\text{Mean}-2\sigma, \text{Mean}+2\sigma] = [22.016-2*4.797, 22.016+2*4.797] = [12.422, 31.610]$

Compare two sample's key parameters:

Test Method	Mode	Mean	Median	Sd	95% CI
Congruent	10-14. 999	14. 05	14. 36	3. 56	[6. 93, 21. 14]
Incongruent	20-24. 999	22. 016	21. 02	4. 797	[12.422,31.610]

To compare the congruent and incongruent 's mode value (mode of frequency), mean value and median value, the incongruent test likely increase the spoke time than congruent test. But two sample data have overlapped CI, we can't fully to get the test result is for retaining null hypotheses or rejecting null hypotheses.

Question 5: Perform the statistical test and interpret your results

Because we can't get the population data of the Stroop test, and can't know the standard deviation of population, there only have small sample data. The sample size is 24 which is smaller than 30. And according to Histogram in previous section, the sample data of congruent and incongruent are likely normal distribution. So we choice T test to instead of Z test to check the difference significant between two test.

Now I use one tail dependent T test to get the result for retaining null hypotheses or rejecting null hypotheses. And this is "Repeated Measures Design".

- The Sample size n is 24;
- And Df is 23;
- The $\alpha = 0.05$;

Set data in congruent sample is as X_i and data in incongruent sample is Y_i .

- We calculate the difference as: $Y_i - X_i$;
- Get the mean of difference is: 7.965;
- Get the Sd of the difference is: 4.865 ;
- Then get the SEM is 0.993. The formula is $SEM = Sd/\sqrt{n}$;
- Get T statistic value is 8.02. The formula is $T \text{ statistic} = (\bar{Y} - \bar{X}) / (Sd/\sqrt{n}) = 7.965 / (4.865/\sqrt{25}) = 8.02$;
- The T critical value for one tail 0.05 is 1.714 (Df is 23).

The T statistic value is much bigger than T critical value.

So we can get the conclusion to reject the null hypotheses.

Question 6: Digging deeper and extending the investigation

NULL