Statistics: The Science of Decisions Project Instructions

Thiravia Nandhan P
January 28, 2019

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:

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

   a. **Independent Variable**: The independent variable is the font color / word being the same or different.
   b. **Dependent Variable**: The dependent variable is the time taken to identify the Congruent or Incongruent words.

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

   In here the same set of people were first tested for Congruent words and then the same set of people were tested for Incongruent words. As per the data sheets there is no mention of Population Mean / Standard Deviation, hence t-tests has to be performed, because z-tests are used when we have the Population Mean / Standard Deviation information provided.

   Note: This problem looks exactly similar to Lesson 26: L10a-t-tests Part 1: Quiz- Keyboard

   Hypothesis:
   Going with the t-tests, following is my hypothesis.

   a. **H0 (Null Hypothesis)**: There is no significant difference in the Time taken to identify Congruent or Incongruent words, i.e. the average population mean for the Time taken to identify the colour of the words in Congruent and Incongruent condition would be same.
   $\mu I = \mu C$

   b. **Ha (Alternative Hypothesis)**: There will be a significant difference in the Time taken to identify the Congruent and Incongruent words, i.e. the average population mean for the Time taken to identify the colour of the words in Congruent and Incongruent condition would be significantly different.
   $\mu I \neq \mu C$

   Our goal is the prove the Alternative Hypothesis.

Symbol Description:

μC: Mean of Time taken to identify Congruent words

μI: Mean of Time taken to identify Incongruent words

Statistical Test:

The hypothesis mentions that the time taken to identify Incongruent words would be not equal to that of for Congruent words, hence it would be a:

a) both-directional
b) two-tailed
c) paired
d) Dependent t-test

because smaller set of people (< 30) are undergoing 2 conditions and we are un-aware of the Population parameters such as Population Mean and Standard Deviation, which otherwise would be a Z- test.

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

    a. **Congruent Identification:**
        i.   Mean: 14.051125
        ii. Median: 14.3565
   iii. Standard Deviation: 3.559357958
    b. **Incongruent Identification:**
        i.   Mean: 22.01591667
        ii. Median: 21.0175
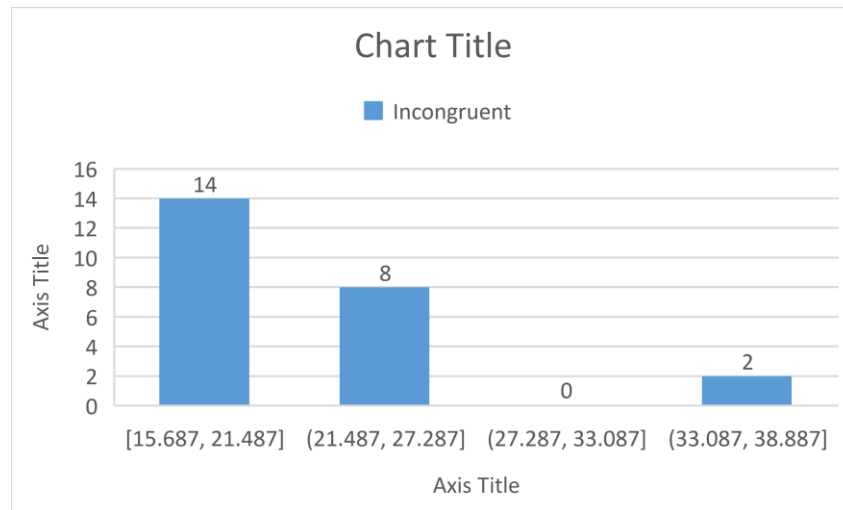   iii. Standard Deviation: 4.797057122
    c. Other calculations are presented below in Q #5.

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.**
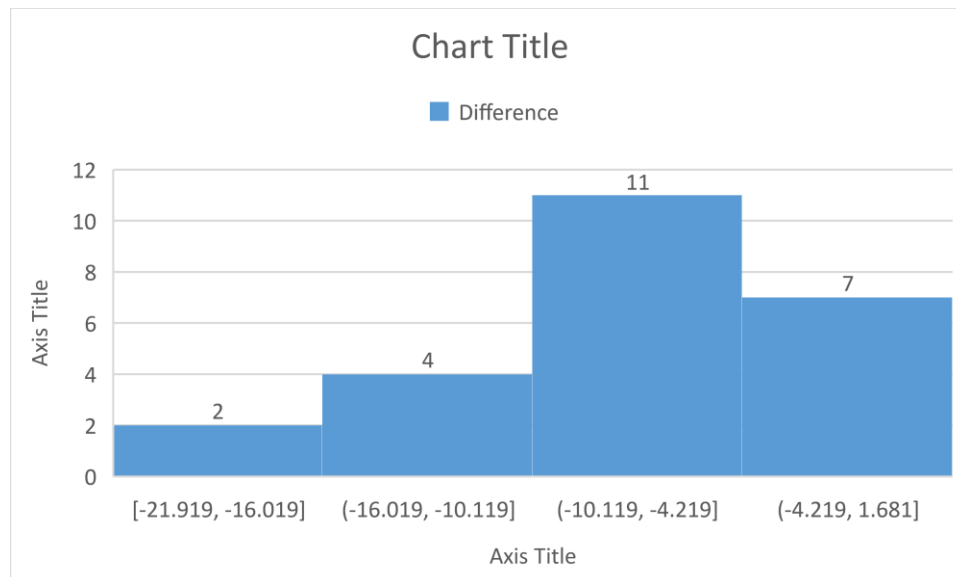
    a. Below is the Histogram for Congruent data. The Histogram looks to be a Positively Skewed distribution, with maximum people identifying Congruent words in faster time.
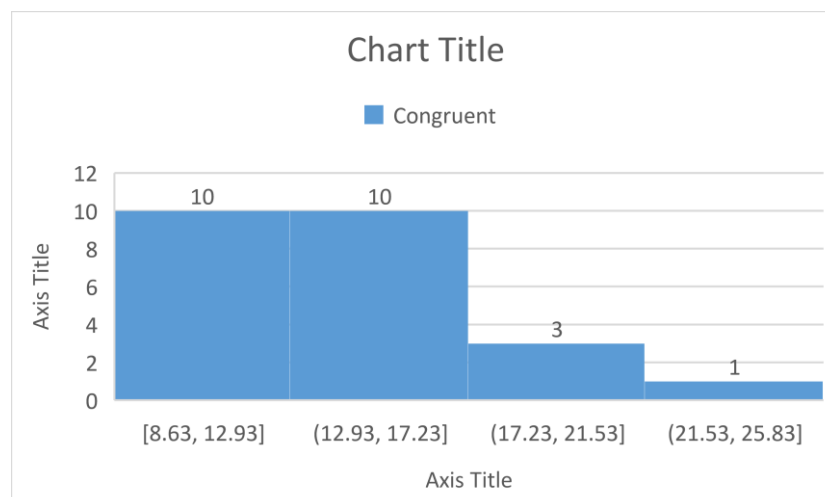
b. Below is the Histogram for Incongruent data. The Histogram also looks to be a Positively Skewed distribution, with maximum people identifying Incongruent words in faster time.

Chart Title

Incongruent

Bin	Frequency
[15.687, 21.487]	14
(21.487, 27.287]	8
(27.287, 33.087]	0
(33.087, 38.887]	2

Axis Title (y): Axis Title
Axis Title (x): Axis Title

c. Below is the Histogram for Difference between Congruent and Incongruent time values. It is a negatively skewed distribution.

Chart Title

Difference

Bin	Frequency
[-21.919, -16.019]	2
(-16.019, -10.119]	4
(-10.119, -4.219]	11
(-4.219, 1.681]	7

Axis Title (y): Axis Title
Axis Title (x): Axis Title

d. Below is the boxplot that shows that Incongruent words identification takes a higher amount of time than the Congruent words identification, but it is still a sample and it cannot be true for the entire population.
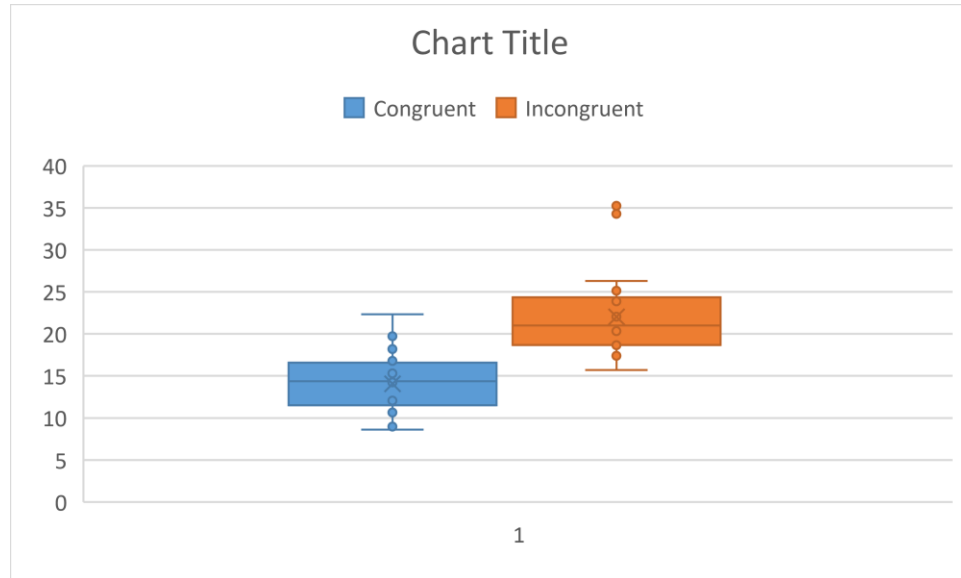
5. **Now, perform the statistical test and report your results. What is your confidence level and your critical statistic value? Do you reject the null hypothesis or fail to reject it? Come to a conclusion in terms of the experiment task. Did the results match up with your expectations?**

Statistical Test:

Congruent	Incongruent	Difference= Congruent - Incongruent	Squared Deviation From Mean
12.079	19.278	-7.199	0.586436877
16.791	18.741	-1.95	36.17771879
9.564	21.214	-11.65	13.58076046
8.63	15.687	-7.057	0.82408571
14.669	22.803	-8.134	0.02863146
12.238	20.878	-8.64	0.455906293
14.692	24.572	-9.88	3.66802296
8.987	17.394	-8.407	0.19554821
9.401	20.762	-11.361	11.53423104
14.48	26.282	-11.802	14.72416779
22.328	24.524	-2.196	33.27895729
15.298	18.644	-3.346	21.33323646
15.073	17.51	-2.437	30.55648071
16.929	20.33	-3.401	20.82819438
18.2	35.255	-17.055	82.63188754
12.13	22.158	-10.028	4.256828627

18.495	25.139	-6.644	1.744490627
10.639	20.429	-9.79	3.33138546
11.344	17.425	-6.081	3.548671043
12.369	34.288	-21.919	194.7199302
12.944	23.894	-10.95	8.911468793
14.233	17.96	-3.727	17.95887821
19.71	22.058	-2.348	31.54834863
16.004	21.157	-5.153	7.906172377

Descriptive Statistics:

Mean	-7.964791667
Variance	23.66654087
Standard Deviation	4.86482691
Standard Error	0.993028635
$t_{statistics}$	-8.020706944

a. **Confidence Level:**
   a. 95 %
   b. **α = 0.05**
   c. $t_{critical}$ **= 2.069**
b. $t_{statistics}$ is beyond the $t_{critical}$ hence we **Reject the (H_0) Null Hypothesis, i.e., there is a significant difference in the mean task completion time between both conditions.**
c. Effect Size Measure:
   a. **Cohen's d**: -1.637219949
   b. The P-Value is < .00001.
d. Confidence Interval of 95% is as below:
   a. **CI:** -10.01936791 to -5.910215421
   b. This means people identify Congruent words faster than Incongruent words.

Note: The calculation Spreadsheet is also attached for further reference.

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!**

Practice makes a man perfect.

If every person practices identifying Incongruent words, then definitely the person would take same or even lesser time to identify the Incongruent words. It is just a matter of Practice and nothing else.
A person in his daily life come across a lot of words, but when it comes to color he come across a very little time and that too with a word that mentions a color but written with a different font color.
This effect can be reversed or equaled if we identify a population sample who have been trained to identify Incongruent words.

7. **Reference Section**

   a) https://en.wikipedia.org/wiki/Stroop_effect
   b) https://keydifferences.com/difference-between-t-test-and-z-test.html
   c) https://www.socscistatistics.com/pvalues/tdistribution.aspx
   d) http://statisticsbyjim.com/basics/measures-central-tendency-mean-median-mode/
   e) https://study.com/academy/lesson/z-test-t-test-similarities-differences.html
   f) https://www.tableau.com/sites/default/files/media/which_chart_v6_final_0.pdf