The Science of Decisions

Kris Anaya

February 13, 2017

Abstract

*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 each case, we measure the time **t**, it takes to name the ink colors in equally-sized lists. Each participant will go through and record a time from each condition.*

1 Introduction

In this task, we are asked to measure the time it takes to name the ink colors in equally-sized lists. In the *congruent* words, the words that are displayed are color words whose names match the colors in which they are printed. In the *incongruent* words conditions, the words displayed are color words whose names do no match the colors in which they are printed. Meaning, what is the probability it takes our independent variable t to name the colors in the *congruent* and *incongruent* words within these two independent samples which are dependent on t:

$$P(X = congruent \mid Y = incongruent) \tag{1}$$

Within this project we test that there is no significant change in speed between the *congruent* and *incongruent* samples. However, this will be tested against alternative changes between our *congruent* and *incongruent* samples given our independent variable t at $\alpha = 0.05$:

$$\begin{aligned} H_0 &: \mu_c - \mu_I = 0 \\ H_A &: \mu_c - \mu_I \neq 0 \end{aligned} \tag{2}$$

We will perform these task by using a Pythonic language module which we have created. This module is aptly named *statistics.py*. The resources for our code may be found within our GitHub page *http://github.com/krismanaya* under *UdactiyDataScience*. Within our methods, we may conclude that a Type I error is very unlikely. Therefore, the test procedure will be based on a rejection H_0 at an α level of 0.05. In conclusion, we will show there is a significant difference between naming *congruent* and *incongruent* colors by relying on our independent variable t.

2 Methods

2.1 sample

The sample population of interest is taken from the Stroop dataset provide by the Udacity Class. This data set may be viewed on Udacity's Google drive website. The sample population that is provided consists of a list of patients who were tested twice, one saying the words with applying the *congruent* method and then tested again saying the word with the incongruent method. Both samples were done under t constraint.

2.2 design

A repeated measures longitude two-test randomized experimental design was provided for this study. We wish to analyze the effect size of the given treatments. The given variables and measures will be the mean, size, standard-error, t-statistics, confidence interval; degrees of freedom, t-critical and Bessel's approximation.These measures and variables will be denoted as:

Stroop variables:

1. $\bar{X}_c :=$ congruent mean
2. $\bar{X}_I :=$ incongruent mean
3. $N_{c,i} :=$ size of list
4. $\alpha :=$ alpha level
5. $t_c :=$ t-critical interval
6. $t_{stat} :=$ t-statistic
7. $df :=$ degrees of freedom
8. $SE :=$ standard error
9. $\sigma_{c,i} :=$ standard deviation (Bessel)
10. $CL :=$ confidence level

Within this design we test that there is no significant change in speed between the *congruent* and *incongruent* samples. However, this study will be tested against alternative changes between our H_0 given our independent variable t at $\alpha = 0.05$:

$$H_0 : \mu_c - \mu_I = 0$$
$$H_A : \mu_c - \mu_I \neq 0 \tag{3}$$

3 Results

3.1 variability

Within the results section we round all data points two decimals places. We begin by locating the variables, we used our Pythonic module *statistics.py* to create a two-list aptly named congruent and *incongruent*. Secondly, we calculated the mean $\bar{X}_{c,i}$ and $\sigma_{c,i}$ with help of our functions:

Table 1: Stroop statistics

Variable	$\bar{X}$	σ
Congruent	14.05	3.56
Incongruent	22.02	4.80
n		24

```
congruent = make_a_list()
incongruent = make_a_list()
x_c = mean(congruent)
>>> 14.05
x_i = mean(incongruent)
>>> 22.02
sigma_c = bessels_correction(congruent)
>>> 3.56
sigma_i = bessels_correction(incongruent)
>>> 4.80
```

The reader may have observed that within the mean and standard deviation, the length and average t is much smaller for our congruent list of data points. This may suggest to us that the first test has an overall faster time of saying words.

3.2 central tendencies

Within our central tendencies we wish to calculate the standard error, t-statistic, degrees of freedom and t-critical at $\alpha = 0.05$:
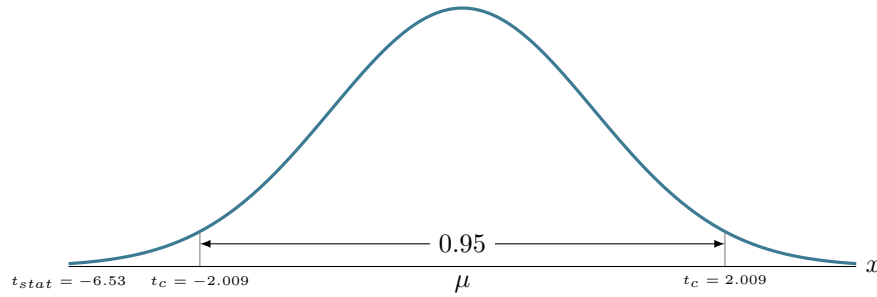
Table 2: Stroop statistics

Central Tendencies	
SE	1.22
df	46
t_{stat}	-6.53
t_c	2.009
α	0.05
CL	.95

```
SE = independent_standard_error(congruent,incongruent,len(congruent),len(incongruent))
>>> 1.22
df = degrees_of_freedom_indy(congruent,incongruent)
>>>  ((24, 24), 46)
t_stat = t_stat_indy(x_c,x_i,0,0,sigma_c,sigma_i,24,24)
>>> -6.53
# t_c = use chart*
```

Here we visualize a t-distribution at α level denoted above and at 0.95 confidence interval. As we can see our t_{stat} variable is well below the critical value. This suggests that it is possible there is some significant changes between the *congruent* and *incongruent*. Because of this we shall reject H_0 and conclude that there exist a probability $P < \alpha$.

4   Discussion

The effects responsible for a rejection of H_0 could be due to the patient's eyes becoming comfortable with words and colors that when they are ready to observe the *incongruent* test their eyes have still not settled. One test you could do is switch the test and perform a Stroop task of *incongruent* then *congruent* and see if the samples differ. Another similar approach for the test would make the patient perform the task multiple times to see if there is some level of improvement after practicing the *incongruent* tests. Finally, a last possibility could be age and eyesight. Do these patients are different in age are they similar? What are the patients eyesight, is it normal or do some patients struggle with color? We do not have more information within the data set. The only information given to us is the independent variable t.