T-Testing a Perceptual Phenomenon – The Stroop Effect

Udacity Data Analyst Nanodegree: Inferential Statistics Project: GitHub

Author: Jubin Soni

Date: September 17, 2017

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. To try out the Stroop task or learn more about it, go to this link, which has a Java-based applet for performing the Stroop task and record the time you received on the task.

Project Overview

In this project, I have investigated a classic phenomenon from experimental psychology called the Stroop Effect. I have learned a little bit about the experiment, created a hypothesis regarding the outcome of the task, then have gone through the task. I have looked at the data collected from others who have performed the same task and have computed some statistics describing the results. Finally, I have interpreted the results in terms of hypotheses.

Data download: The Stroop Data

Why this Project is useful?

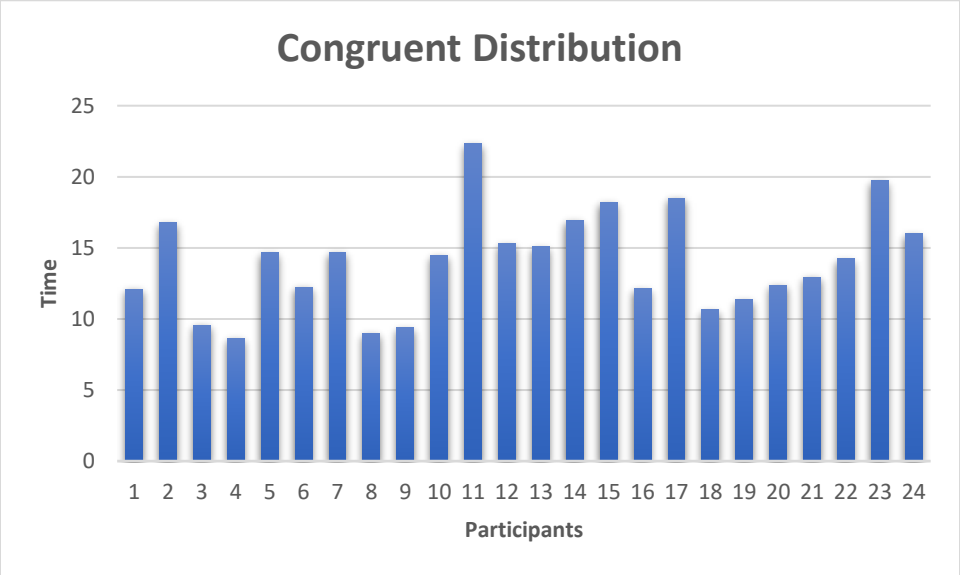
Statistics is a major component of data analysis, it allows us to investigate data and make inferences based on our observations. In this project, I have learned:

- How to identify components of an experiment.
- How to use descriptive and inferential statistics to describe qualities of a sample.
- How to perform P-tests, set up a hypothesis test, make inferences from a sample, and draw conclusions based on the results.

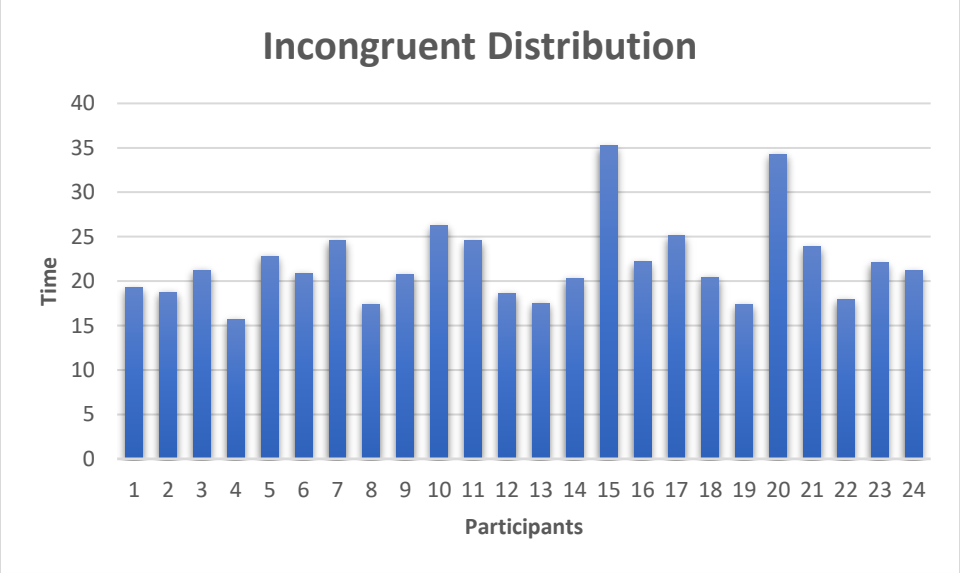
Exploring the dataset

The data has congruent and incongruent distributions as explained in the background information section. The sample size is small n=24 and standard deviation is not known.

Bar graph of congruent distribution representing time taken by each participant.



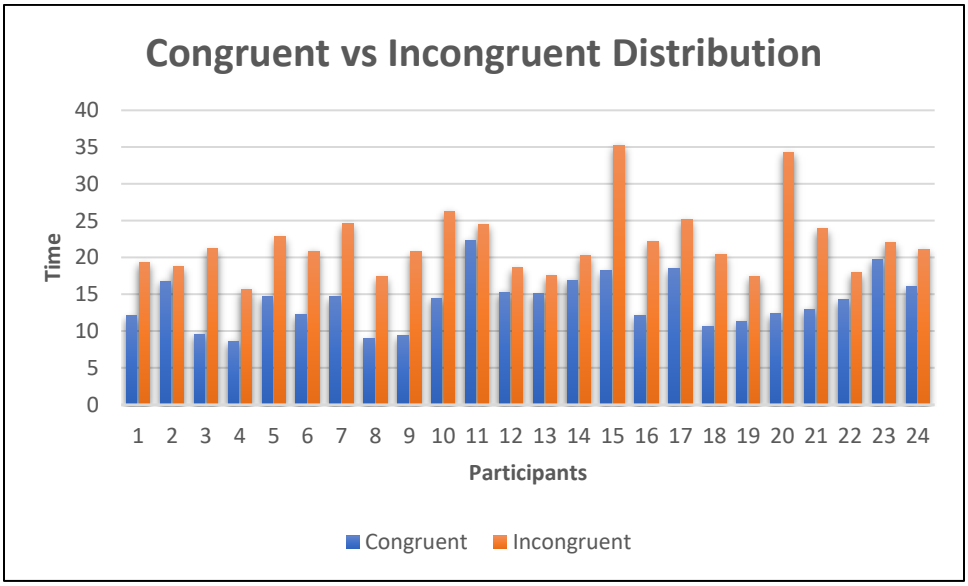
Bar graph of incongruent distribution representing time taken by each participant.



Questions Answered for the Investigation:

1. What is our independent variable? What is our dependent variable?
   Independent variable: the words condition (congruent or incongruent words).
   Dependent variable: the time it takes to name the ink colors in equally-sized lists.

2. What is an appropriate set of hypotheses for this task? What kind of statistical test do you expect to perform? Justify your choices.
   Let us compare congruent and incongruent distributions:



We can see that the times taken by participants for incongruent is higher than the congruent distributions. Therefore, the null hypotheses should be that the mean time for the color recognition for congruent words $(\bar{x}_c)$is equal to or greater than the mean time for incongruent words $(\bar{x}_i)$, therefore implying a one-tailed test. The alternate hypotheses should be that the congruent words mean is less than the incongruent words mean.

$$H_0: \bar{x}_c \geq \bar{x}_i$$

$$H_A: \bar{x}_c < \bar{x}_i$$

μ is the population mean, *c* represents the congruent words condition, and *i* represents the incongruent words condition.

We will perform t-test because:

- The population variance or standard deviation is unknown.
- The sample size $n = 24$ (less than 30). When the sample size is less than 30, the sample data no longer approximates a normal distribution which makes use of a Z-value inappropriate.

The following assumptions are required for t-test for dependent means:

- Interval or ratio scale of measurement (approximately interval).
- Random sampling from a defined population.
- Samples or sets of data used to produce the difference scores are linked in the population through repeated measurement, natural association, or matching.
- Scores are normally distributed in the population; difference scores are normally distributed.

The t-test we perform will be one-tailed because:

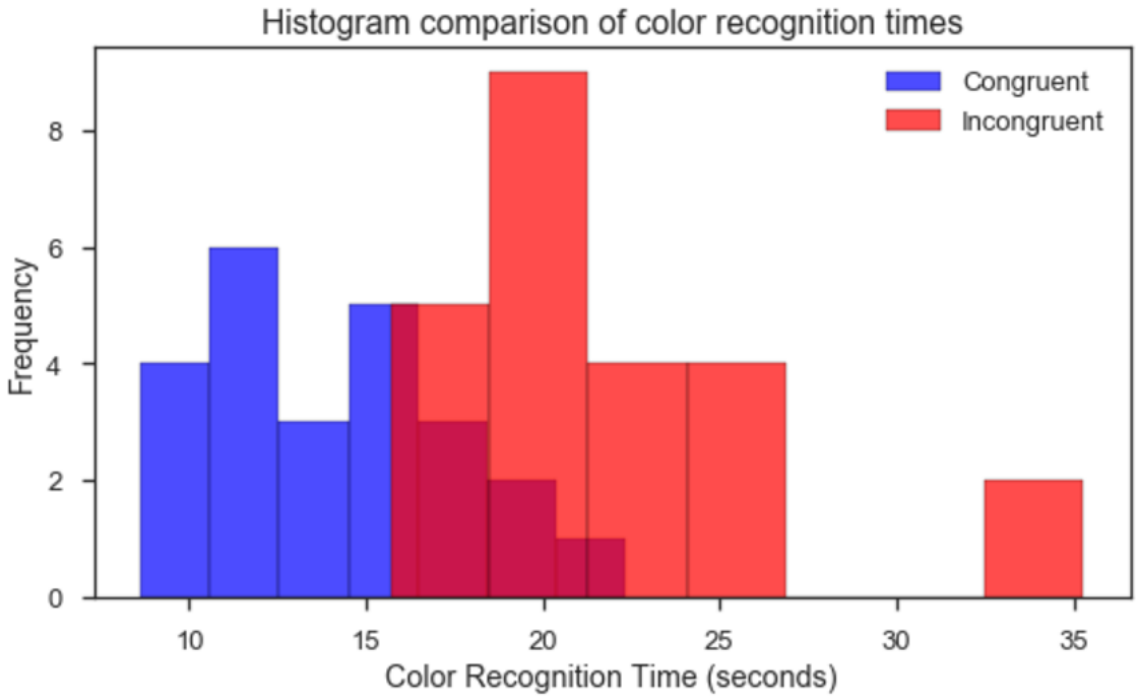
- The assumption that incongruent word conditions will not improve recognition times.
- One-tailed test allows for a more scrutiny examination of negative impact word conditions on recognition times.

The t-test will be of dependent sample variety because the same subject is exposed to two conditions and tested for each, which are defining criteria for "within-subjects" or "repeated-measures" statistical tests.

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

Statistic	**Congruent**	**Incongruent**	**Difference (C-I)**
Sample size (n)	24	24	24
Mean ($\bar{x}$)	14.051125	22.01591667	-7.964791667
Median	14.3565	21.0175	7.6665
Variance	12.66902907	23.01175704	23.66654087
Standard Deviation (σ)	3.559357958	4.797057122	4.86482691
Standard Error (SE)	0.726550901	0.979195185	0.993028635
Min	8.63	15.687	1.95
Max	22.328	35.255	21.919

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.



The bucket range of the time for congruent words data is 8 seconds to 23 seconds. While the bucket range of time for incongruent words data is 16 seconds to 36 seconds. Incongruent is distributed noticeably at higher right to congruent distribution.

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?

$\alpha = 0.05$

$df = 23$

$t_{crit} = -1.714$ (t-table)

Since our sample size is same we will not do pooled-standard error, instead we will calculate the standard error using below formula:

$$SE = \frac{\sigma}{\sqrt{n}} = \frac{4.8648691}{\sqrt{24}}$$

$$SE = 0.9930372$$

$$t = \frac{\bar{x}_c - \bar{x}_i}{SE} = \frac{-7.964791667}{0.9930372}$$

$t = -8.0206$

$P\ value = 0.1000$

$Margin\ of\ Error\ (ME) = 2.069 * 0.9930372 = 2.054$

$Confidence\ Interval$:

$lower\ limit = -7.964791667 - 2.054 = -10.01$

$upper\ limit = -7.964791667 + 2.054 = -5.91$

At 95% confidence interval ($\alpha = .05$) and 23 degrees of freedom, the critical statistic value for one-tailed test in the negative direction is -1.714. The calculated t-statistic for the difference in color recognition times is -8.0206. Since the t-statistic is in the critical region, we reject the null hypothesis (H_0: $\bar{x}_c \geq \bar{x}_i$).

This result is not quite statistically significant as $P\ value \leq 0.1$.
This result is meaningful because it is very unlikely that 7.964 seconds time difference in congruent and incongruent mean is obtained, if the two means are actually same.

We can conclude that the congruent condition takes less time to name ink colors than incongruent condition.

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!

My hypotheses for this effect is that our brain focuses more on reading the word rather than recognizing a color when eyes are presented with a colored word. To recognize a color, one has to override the brains natural tendency of reading the word first. This override takes time and is likely not always successful, which means re-analyzing and recognizing the error, which costs more time.
An alternative task to Stroop Effect will be to use your non-dominant hand for routine chores. I am right-handed and I noticed that when I try to brush my teeth with my left hand it takes more mental effort.

Sources

1. Wikipedia: Numerical Stroop Effect
2. GraphPad: P value calculations
3. GFC: Benefits of using your opposite hand
4. Yale: T-tests of significance