Color Words and Ink Colors

Abdulaziz M. Alqumayzi

(DS-510) – Discovering Statistics Using R

Colorado State University – Global Campus

Dr. Hasan Aljabbouli

November 13, 2020

Color Words and Ink Colors

Introduction

We will gather data from involvement in a measurement task in this assignment activity, establish a hypothesis about the outcome, test the hypothesis, and summarize the findings of the study.

Scenario

A well-known phenomenon studied in psychology is the Stroop Effect. Lists of terms shown in various colors are shown to the participants in the experiment. The person is encouraged to say the color of the ink in which the word is printed out loud. The ink color of the term RED, for instance, is blue. Two sections compose of the experiment. The color of the ink and the meaning of the term are congruent throughout the first part (they fit, for instance, RED). They do not fit in the second section (e.g.: the word is PURPLE, and the color is ORANGE).

Part I: Data Exploration

Firstly, "stroop data 1 .csv" file was downloaded and imported as in figure 1, containing data gathered for various participants in the trial, which is organized into two columns. The first column indicates the time a subject needed to finish the congruent part of the experiment, and the second, the incongruent part of the experiment.

Figure 1

import the data to Rstudio environment

```
stroop_data_1_ <- read_csv("~/Master DS/DS510/Critical Thinking/CT4/stroop_data _1_.csv")
```

Secondly, descriptive statistics on this dataset, including measures of central trend and variability, will be shown in figure 2. A summary() function used to summarize the dataset, we can see the minimum, first quartile, median, mean, third quartile, and maximum values. Also, the figure 2 shows the variance and standard deviation using var() and sqrt() functions.

The variance of congruent and incongruent is 12.669 and 23.011 respectively, as the standard

deviation of 3.559 and 4.797.

Figure 2

functions and outputs of the measure's central tendency and variability codes

```
> summary(stroop_data_1_)
   Congruent         Incongruent
 Min.   : 8.63   Min.    :15.69
 1st Qu.:11.90   1st Qu.:18.72
 Median :14.36   Median :21.02
 Mean   :14.05   Mean    :22.02
 3rd Qu.:16.20   3rd Qu.:24.05
 Max.   :22.33   Max.    :35.26
> var(stroop_data_1_)
              Congruent Incongruent
Congruent     12.669029    6.007123
Incongruent    6.007123   23.011757
> sqrt(var(stroop_data_1_))
              Congruent Incongruent
Congruent      3.559358    2.450943
Incongruent    2.450943    4.797057
```

Part II: Hypotheses

In this test, the statement of the hypothesis is that the null hypothesis means the

congruent is equal to the incongruent time, the alternative hypothesis means the congruent

time is greater than the incongruent. The hypothesis statement will be:

The null hypothesis **H0**: $\mu d = 0$

The alternative hypothesis **Ha**: $\mu d > 0$

The alternative hypothesis **Ha**: $\mu d < 0$

The alternative hypothesis **Ha**: $\mu d \neq 0$

We are comparing two different results of each experiment taker to see how much it

take the two tests and compares between them if the two tests are taking the same amount of

time or less or more? So, this test called: Comparing Means of Dependent Samples. Because

the two tests related to the same person, but with two experiments to find out the time taken

of each participant. luckily, R has simplified the steps to us with only one code that shown in

the following figure 3. t.test() function does the calculation of the test statistics, the degree of

freedom, p-value, by default the confidence interval is 0.95 and many more information

depend on which test you want to do.

Figure 3

t.test function code and output

```
> t.test(stroop_data_1_$Congruent,stroop_data_1_$Incongruent, paired = TRUE)

        Paired t-test

data:  stroop_data_1_$Congruent and stroop_data_1_$Incongruent
t = -8.0207, df = 23, p-value = 4.103e-08
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -10.019028  -5.910555
sample estimates:
mean of the differences
          -7.964792
```

Part III: Reporting

In the previous figure 3, the output of the results show us the test statistic is -8.02,

this negative value indicates that the congruent has a smaller mean than incongruent. The

degree of freedom is 23, and the p-value is approximately 0, due to e-08. P-value is smaller

than the significant level 0.05, which is mean we reject the null hypothesis, and this

extremely significant that the null hypothesis means not equal to the alternative hypothesis

mean, as it has shown the figure 3 that the true difference in the mean is not equal to 0.

The confidence interval informs us of the limits that are likely to lie within the true

mean difference. The real mean difference lies between -10.01 and -5.91, we may say. The

meaning of this interval is that it does not include zero (i.e., all limits are negative) so it tells

us that it is doubtful that the true value of the mean difference is zero. In the following figure

4, we can see these boundaries.

Figure 4

congruent, incongruent and difference means boxplot
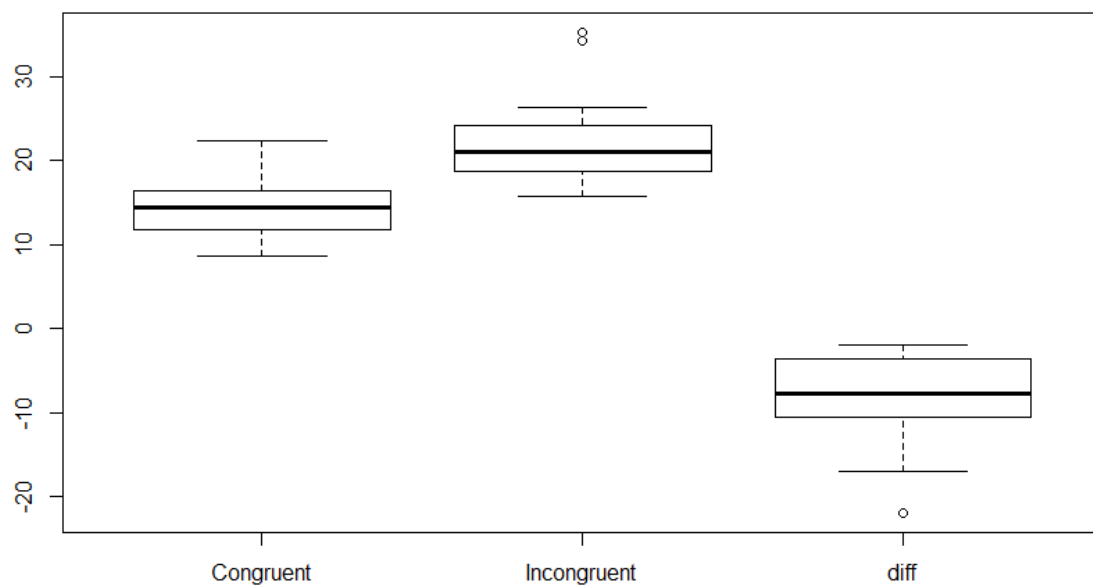
When it occurs at various points in time from the same person, paired samples. In this activity, each participant did two experiments to test the Stroop Effect. The p-value shows us that the two means are significantly different, due to the p-value < 0.05. This sample tells us that the incongruent experiment always will lie in the negative area (as it is shown in the figure 4), which means most likely the population will lies in the incongruent experiment in the negative area, this doesn't mean that our influence is significant in realistic terms. Also, the standard errors must be done.

References

Field, A. P., Miles, J., & Field, Z. (2014). *Discovering statistics using R*. London, UK: Sage.

Peck, R., Short, T., & Olsen, C. (2020). *Introduction to statistics and data analysis*. Boston, MA: Cengage.