Project1_StroopEffect

Max

October 14, 2015

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.

Load libraries

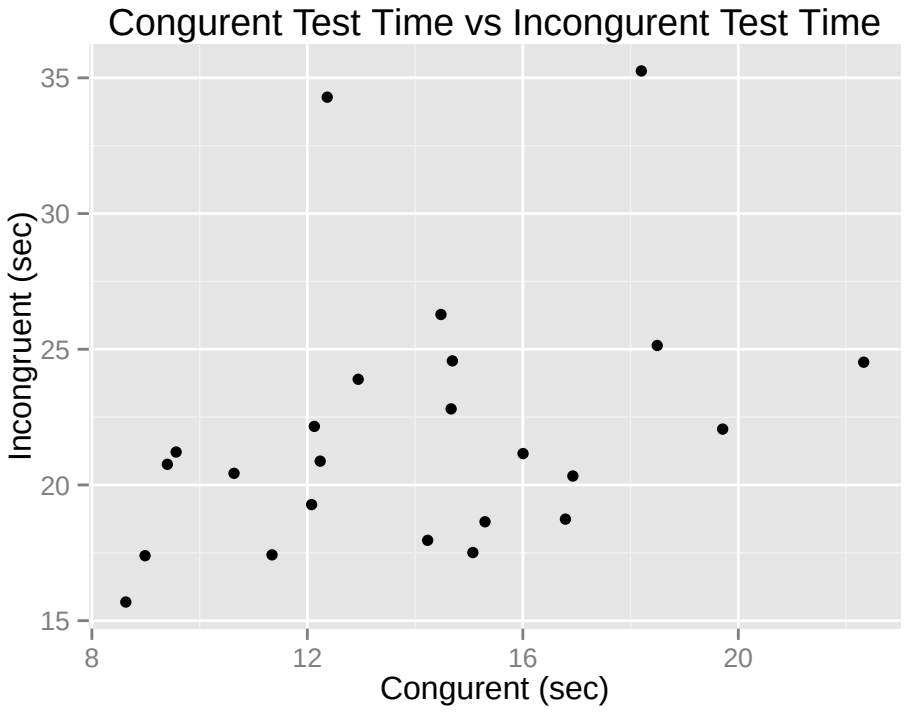
```
library(ggplot2) # For creating beautiful plots
library(knitr)
library(gridExtra) # for ggplot2 stacking plots
```

Load Data

```
stroop <- read.csv("C:/Users/MAX/Documents/Udacity/Data Analyst Nanodegree/Project1_StroopEffect/stroop
```

Explore Data

```
ggplot(data = stroop, aes(x=Congruent, y = Incongruent)) + geom_point() + labs(title = "Congurent Test
```



Project Questions

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

The indpenendent variable is variable is the word/color combination and the dependent variable is the time to complete the test. The researcher changed the word/color combination from congruent to incongruent and measured the time to complete the test for each case.

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

The data collected provides two seperate samples with the same participants for each sample. Additionally, the sample size is less than 30 and the population standard deviation is unknown. Therefore, a paired T-test is appropriate to determine if the treatment (Incongruent words/colors) has an effect on the test time. By performing this T-test, we are assuming the population distribution to be normal and this is reasonable assumption given the sample data appears to be gaussian (see plots in question 4). Because we would expect the stroop effect to increase the time it takes an individual to complete the test, a one-tailed T-test with the following hypothesis is formulated:

- H0 (null hypothesis): mu_c = mu_i
- H1 (alternative hypothesis): mu_c < mu_i

Where; `mu_c` is the population mean for the congruent test and `mu_i` is the population mean for the Incongruent test.

or equivalently:

- H0 (null hypothesis): mu_diff = 0
- H1 (alternative hypothesis): mu_diff < 0

Where; `mu_diff` is poplulation mean of the difference between `mu_c` and `mu_i`.

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

The following provides a summary of the data showing the Min, Max, Mean, Median, and Quartiles for the Congruent test data and Incongruent test data. The Mean and Median provide measures of central tendency. The mean for `Congruent` test data is 14.05 and the mean for `Incongruent` test data is 22.02.

```
stroop_sum <- summary(stroop)
kable(stroop_sum)
```

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

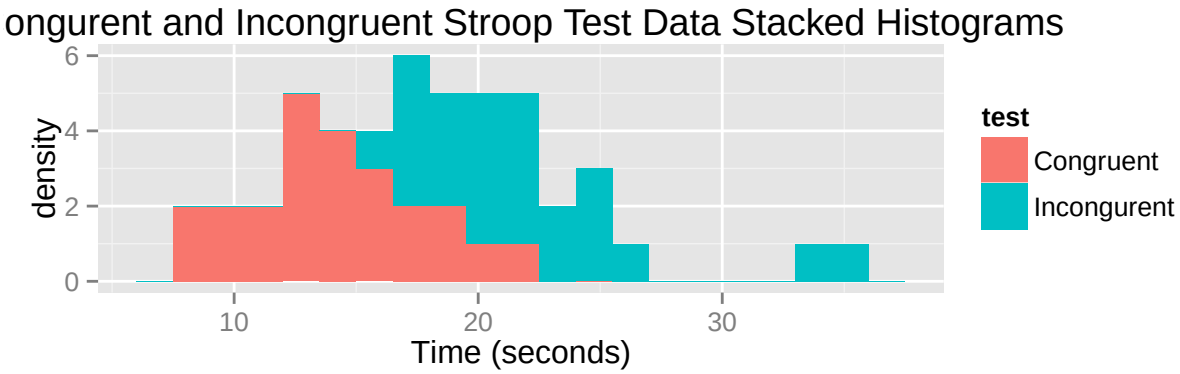
To measure variability, the `sd` function in R is useful. The following code calculates the standard deviation for Congruent and Incongruent test performances for the sample data.

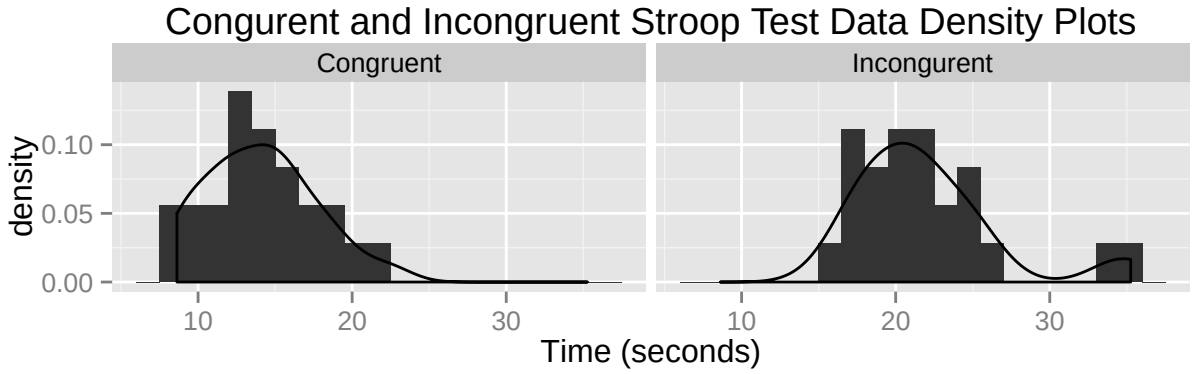
```
con_sd <- round(sd(stroop$Congruent),2)
incon_sd <- round(sd(stroop$Incongruent),2)
```

The standard deviation for `Congruent` test data is 3.56 and the standard deviation for `Incongruent` test data is 4.8. Therefore, this would indicate the incongruent test results have higher variability.

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 following code provides a view of the distributions of the sample `Congruent` and `Incongruent` data. The `Congruent` data distrubition is right skewed (positively skewed). The `Incongruent` data distribution is more normal but has a right tail.





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?

The difference between the congruent test results and incongruent test results is calculated. Then the following information is calculated in order to calculate the T-statistic. * sample mean `mu_diff` * sample standard deviation `s_diff` * sample size `n`.

```
stroop$Difference <- stroop$Congruent - stroop$Incongruent
mu_diff <- mean(stroop$Difference)
s_diff <- sd(stroop$Difference)
n <- dim(stroop)[1]
dof <- n-1
s_diff_manual <- sqrt(sum((stroop$Difference - mu_diff)^2)/(dof)) # manually calculated to confirm stde
t_stat <- round(mu_diff / (s_diff/sqrt(n)),3)
```

At at 95% Confidence level, the T critical values for a one-tailed paired t-test are +/- 1.714. The T-stat is -8.021 which means we reject the null hypothesis. Therefore, the time between the the Incongruent and Congurent test was significantly different and we can say adding the Incongruency (stroop effect) had a causal effect on the time it took to complete the test. These results align with what I expected given the view of the distrubtions in the visualizations generated in question #4.

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!

References

1. https://www.rstudio.com/wp-content/uploads/2015/03/ggplot2-cheatsheet.pdf
2. http://stackoverflow.com/questions/8901330/multiple-histograms-with-ggplot2-position