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.

Questions for Investigation

As a general note, be sure to keep a record of any resources that you use or refer to in the creation of your project. You will need to report your sources as part of the project submission.

1. What is our independent variable? What is our dependent variable?
    - The independent variable is word condition (congruent and incongruent).
    - The dependent variable is time.
2. What is an appropriate set of hypotheses for this task? What kind of statistical test do you expect to perform? Justify your choices.

    H_0: $\mu_{CW} = \mu_{IW}$

    H_A: $\mu_{CW} \neq \mu_{IW}$, where μ_{CW} represents the mean time for congruent words and μ_{IW} represents the mean time incongruent words

    The null hypothesis states that the population mean for the congruent and incongruent test times are equivalent. The alternative hypothesis states that the population mean for the congruent and incongruent test times differ. We will consider the sample data in order to determine whether the two population mean are significantly different.

    We should perform a 2-sided dependent t-test in order to determine whether or not to accept or reject the null hypothesis. We use a t-test because we want to show how different two sample means are from each other. Furthermore, a t-test will be appropriate because of the smaller sample size ($N<30$) and looking at the distribution in the histogram below, the distribution is not highly skewed. Additionally, we do not know the population standard deviation which makes the t-test an appropriate test for

this data set. Lastly, we will use a dependent t-test because there is 2 conditions and the participants are the same for both conditions.

Now it's your chance to try out the Stroop task for yourself. Go to this link, which has a Java-based applet for performing the Stroop task. Record the times that you received on the task (you do not need to submit your times to the site.) Now, download this dataset which contains results from a number of participants in the task. Each row of the dataset contains the performance for one participant, with the first number their results on the congruent task and the second number their performance on the incongruent task.

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

```
> setwd("~/udacity_data_science_nano")
> stroop_data <- read.csv("stroopdata.csv")
> summary(stroop_data)
```

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

VARIANCE

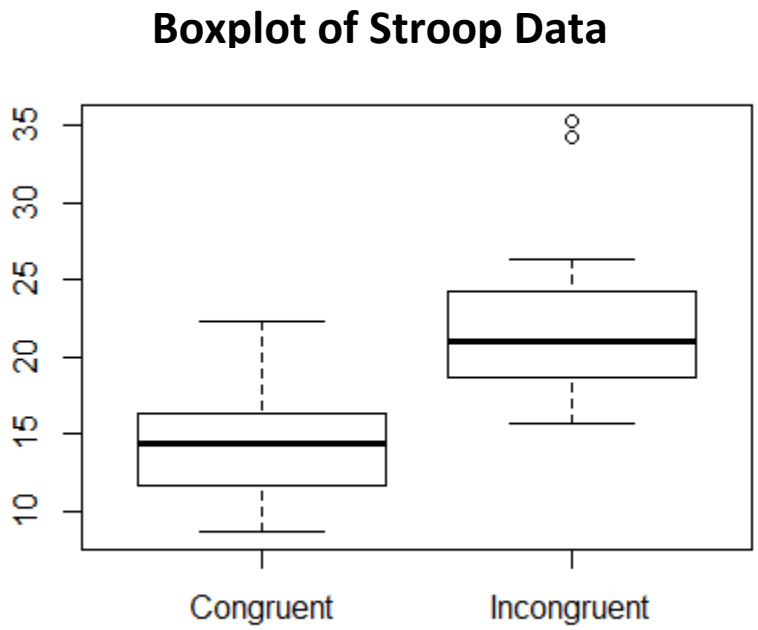
```
> apply(stroop_data,2,var)
  Congruent Incongruent
   12.66903    23.01176
```

STANDARD DEVIATION

```
> apply(stroop_data,2,sd)
  Congruent Incongruent
   3.559358    4.797057
```

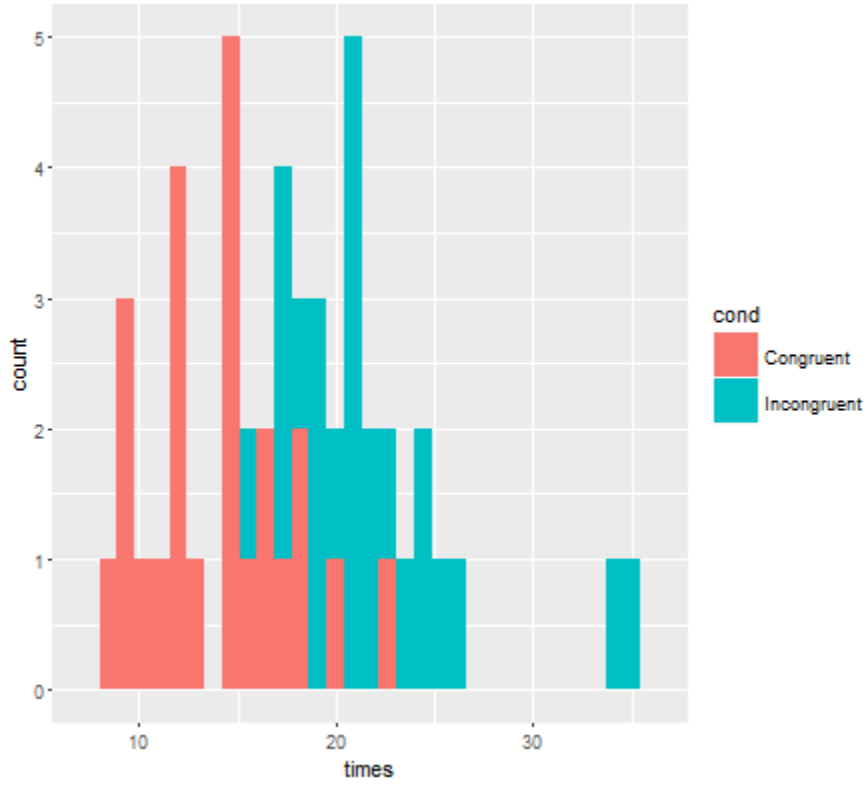
4. Provide one or two visualizations that show the distribution of the sample data. Write one or two sentences noting what you observe about the plot or plots.

```
> boxplot(stroop_data)
```




```
> library(ggplot2)
> Congruent <- stroop_data[1:24,1]
> Incongurent <-stroop_data[1:24,2]
> times <- c(Congruent, Incongurent)
> times_col  <- cbind(times)
> Congruent_title <- rep("Congruent", 24)
> Incongruent_title <- rep("Incongruent", 24)
> cond <- c(Congruent_title, Incongruent_title)
> cond_col <- cbind(cond)
> new_stroop <- data.frame(cond_col, times_col)
> ggplot(new_stroop, aes(x=times, fill=cond))  +  geom_histogram()
```

Histogram of Stroop Data



The range for the congruent data seems to be larger than the incongruent (when ignoring the outliers from the incongruent data). Additionally, the distribution of the times appears to be higher for the incongruent data set.

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

tcritical = ±2.069

Mean difference= -7.9648

tstatistic = $\frac{(\text{mean(a)} - \text{mean(b)})}{(\text{standard\_deviation(a-b)/square\_root(N)})} = \frac{-7.9648}{(4.865/\text{square\_root(24)})} = -8.02$

Confidence level = 95%

Conclusion: We reject the null hypothesis at $p<0.05$. Therefore, there is a significant difference between the means of the incongruent times and the congruent times and the Stroop effect is present. This is in line with what I expected the outcome to be.