P1 Test a Perceptual Phenomenon

Bright Wang

November 6, 2016

Statistics: The Science of Decisions Project - Udacity Data Analyst P1.

Introduction - "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."

Question 1: What is our independent variable? What is our dependent variable?

The independent variable is the difficulty of the word list and if the list is Congruent or incongruent. The dependent variable is the response time.

Question 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 appropriate hypothese for this task is the two-tailed t-test. Two-tailed t-test because we can not make any assumption whether congruent or incongruent is faster.

The Null Hypothesis is $H_0 = \mu_{congruent} - \mu_{incongruent} = 0$ or "There is no difference in the response time between congruent and incongruent samples."

The Alternative Hypothesis is $H_A = \mu_{congruent} - \mu_{incongruent} \neq 0$ or "There is a significant difference in the response time between congruent and incongruent samples."

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

Import CSV and show head

```
df <- read.csv(file="stroopdata.csv", header = T)
head(df)
```

```
##   Congruent Incongruent
## 1    12.079      19.278
## 2    16.791      18.741
## 3     9.564      21.214
## 4     8.630      15.687
## 5    14.669      22.803
## 6    12.238      20.878
```

Congruent Mean:

```
mean(df$Congruent)
```

```
## [1] 14.05113
```

Congruent Median:

```
median(df$Congruent)
```

```
## [1] 14.3565
```

Congruent Standard Deviation:

```
sd(df$Congruent)
```

```
## [1] 3.559358
```

Incongruent Mean:

```
mean(df$Incongruent)
```

```
## [1] 22.01592
```

Incongruent Median:

```
median(df$Incongruent)
```

```
## [1] 21.0175
```

Incongruent Standard Deviation:

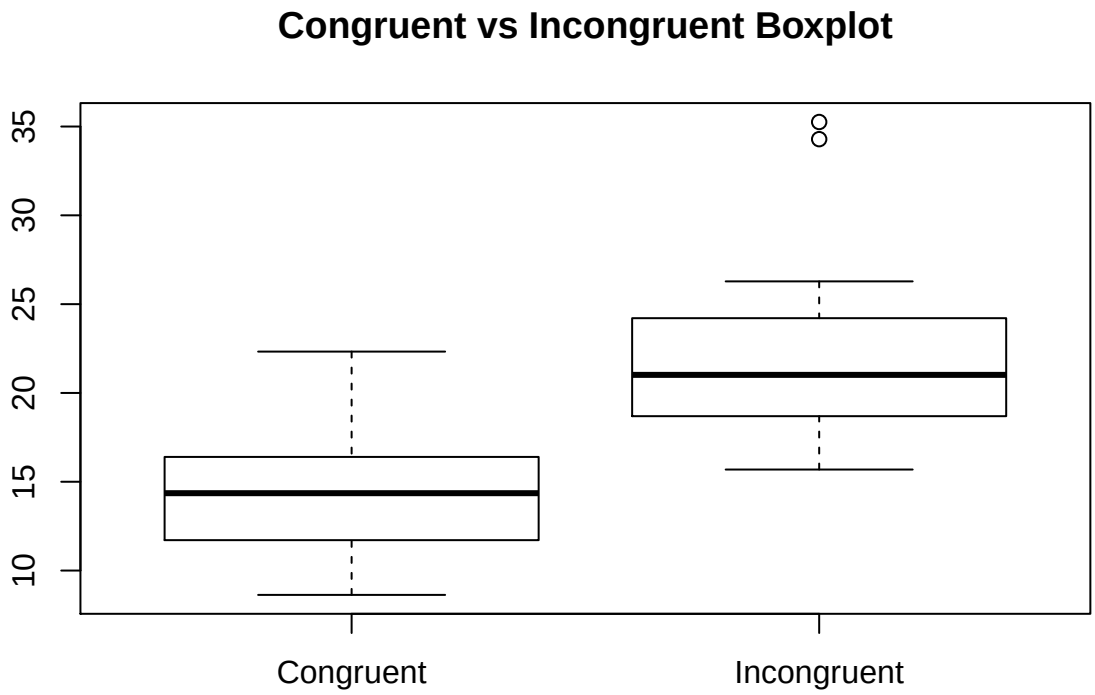
```
sd(df$Incongruent)
```

```
## [1] 4.797057
```

Question 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.

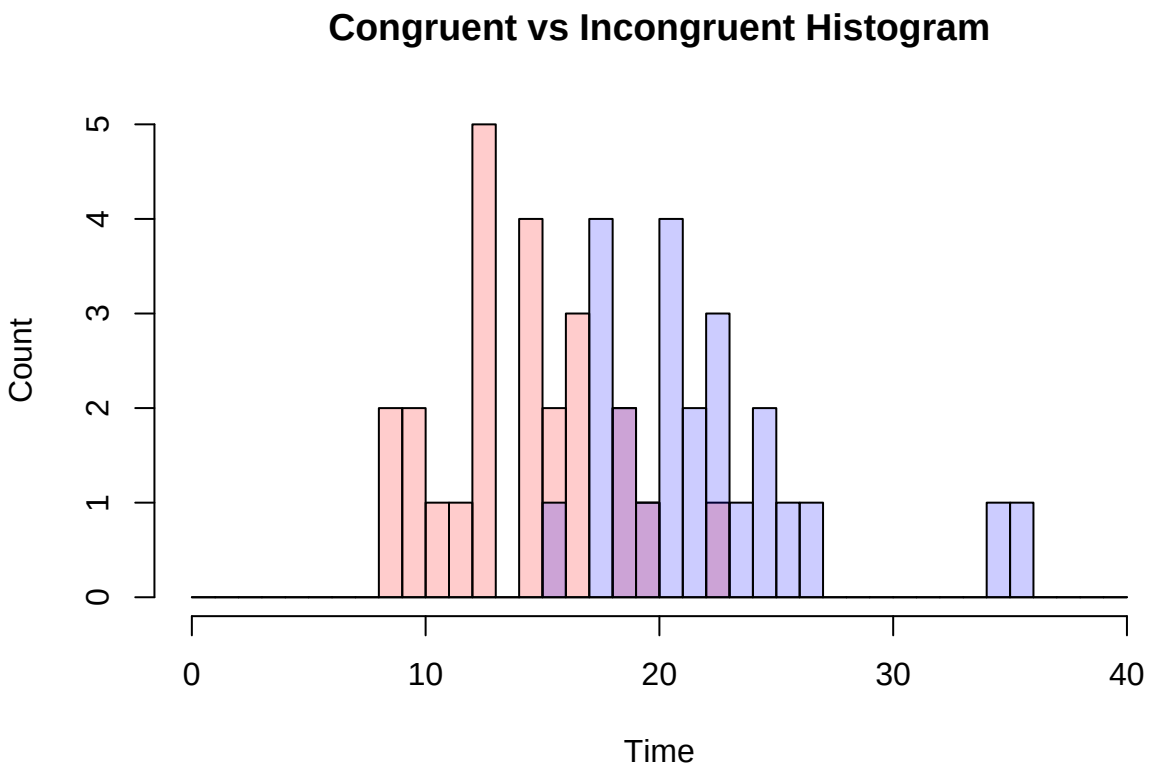
Boxplot of Congruent and Incongruent.

```r
boxplot(df$Congruent, df$Incongruent, names=c("Congruent","Incongruent"),
        main="Congruent vs Incongruent Boxplot")
```



Here is the histogram of Congruent and Incongruent on the same histogram:

```r
hist(df$Congruent, main="Congruent vs Incongruent Histogram", xlab="Time", ylab="Count",
     col=rgb(1,0,0,0.2), breaks = seq(0,40,by=1))
hist(df$Incongruent, col=rgb(0,0,1,0.2), add = T, breaks = seq(0,40,by=1))
```



From these plots, it is a good guess that the data sets would reject the Null Hypothesis as the two data samples seems very different.

Question 4: 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?

```
t.test(df$Congruent, df$Incongruent, alternative = c("two.sided"), paired = T,
       var.equal = T, conf.level = 0.95)
```

```
##
##  Paired t-test
##
## data:  df$Congruent and df$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
```

The confidence level is set at 5%, the t-critical is 2.068, t=-8.02 and the p-value is almost 0 here. The Null Hypothesis is rejected, this matched our expectations from looking at the plots of data.