Test a Perceptual Phenomenon

Udacity : Data Analyst Nanodegree

Trushit Vaishnav

Introduction

In this project, we study a Stroop Task. 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.

Since all the participants will go through both the treatments, this is a case of `dependent samples`. The purpose of the study is to determine if there is a statistically significant difference between average time taken to respond in case of congruent condition and incongruent condition.

RStudio has been used to perform the study.

Following line of code loads the dataset in R environment:

```
stroop<-read.csv("stroopdata1.csv",sep=",")
```

Identification of Variables

We are studying the effect of `color of ink` on the `response time`. Color of ink for the given words leads to two condition: `Congruent` and `Incongruent`.

`Independent Variable`: Color of ink / Condition (congruent or Incongruent)

`Dependent Variable`: Response time

Descriptive Statistics

`Mean` and `Median` are calculated as a measure of central tendency. `Standard deviation` is calculated as a measure of variability. It should be noted that `sd()` calculates `sample standard deviation`.

```
mean_con<-mean(stroop$Congruent)
mean_in<-mean(stroop$Incongruent)
median_con<-median(stroop$Congruent)
median_in<-median(stroop$Incongruent)
std_c<-sd(stroop$Congruent)
std_i<-sd(stroop$Incongruent)
des_con<-c(mean_con,median_con,std_c)
des_in<-c(mean_in,median_in,std_i)
lab<-c("Mean Response Time","Median Response Time","Standard Deviation")
con<-c("Congruent","Incongruent")
result<-data.frame(lab,des_con,des_in)
colnames(result)<-c("Measure","Congruent","Incongruent")
result
```

```
##                Measure Congruent Incongruent
## 1   Mean Response Time 14.051125   22.015917
## 2 Median Response Time 14.356500   21.017500
## 3   Standard Deviation  3.559358    4.797057
```

Visual Exploration of Data

The data has been visually explored using histogram to see the distribution of the data.

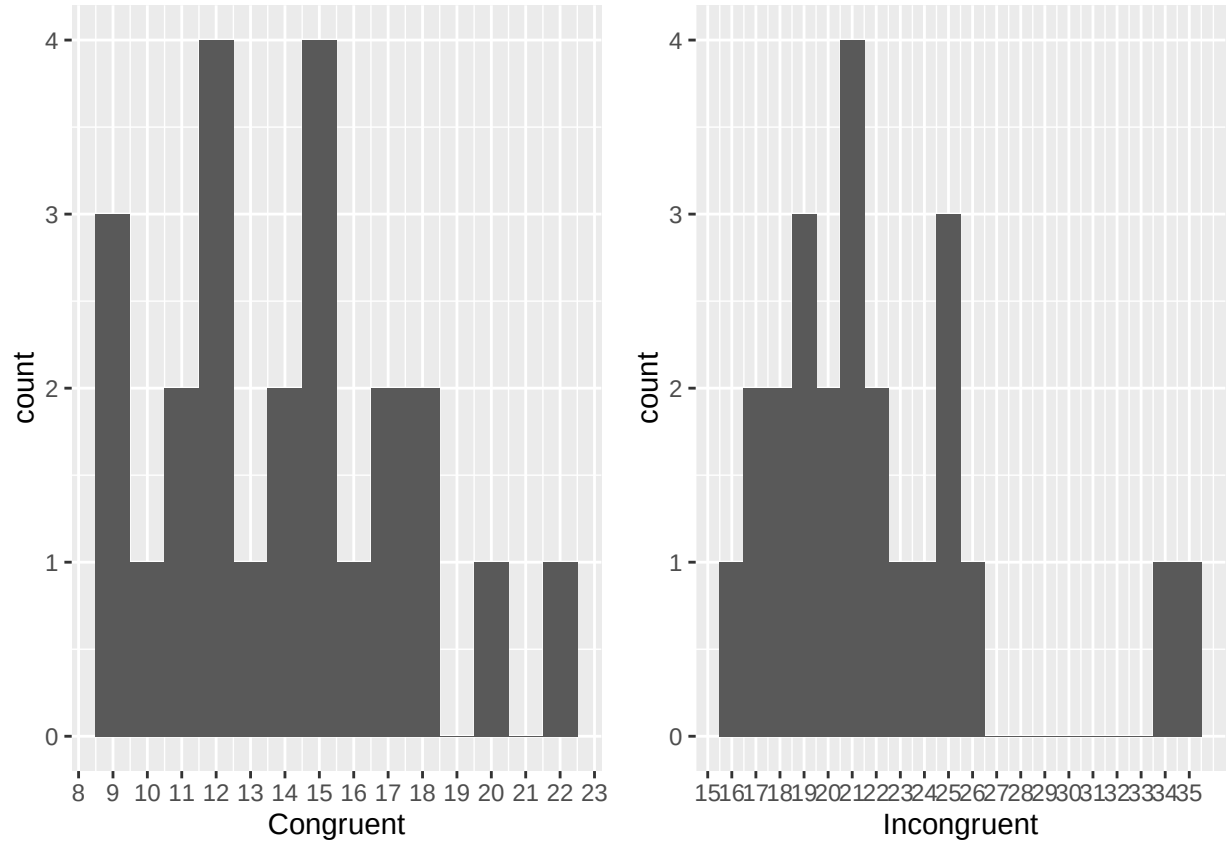
```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.2.5
```

```
library(gridExtra)
```

```
## Warning: package 'gridExtra' was built under R version 3.2.5
```

```
p1<-ggplot(stroop,aes(x=Congruent)) + geom_histogram(binwidth=1)+ scale_x_continuous(breaks = seq(8,25,1
p2<-ggplot(stroop,aes(x=Incongruent)) + geom_histogram(binwidth=1)+ scale_x_continuous(breaks = seq(15,3
grid.arrange(p1,p2,ncol=2)
```



It can be see that data for both the condition, data is NOT normally distributed.

Histogram for congruent condition shows two high bars giving a sense of bimodal like distribution.

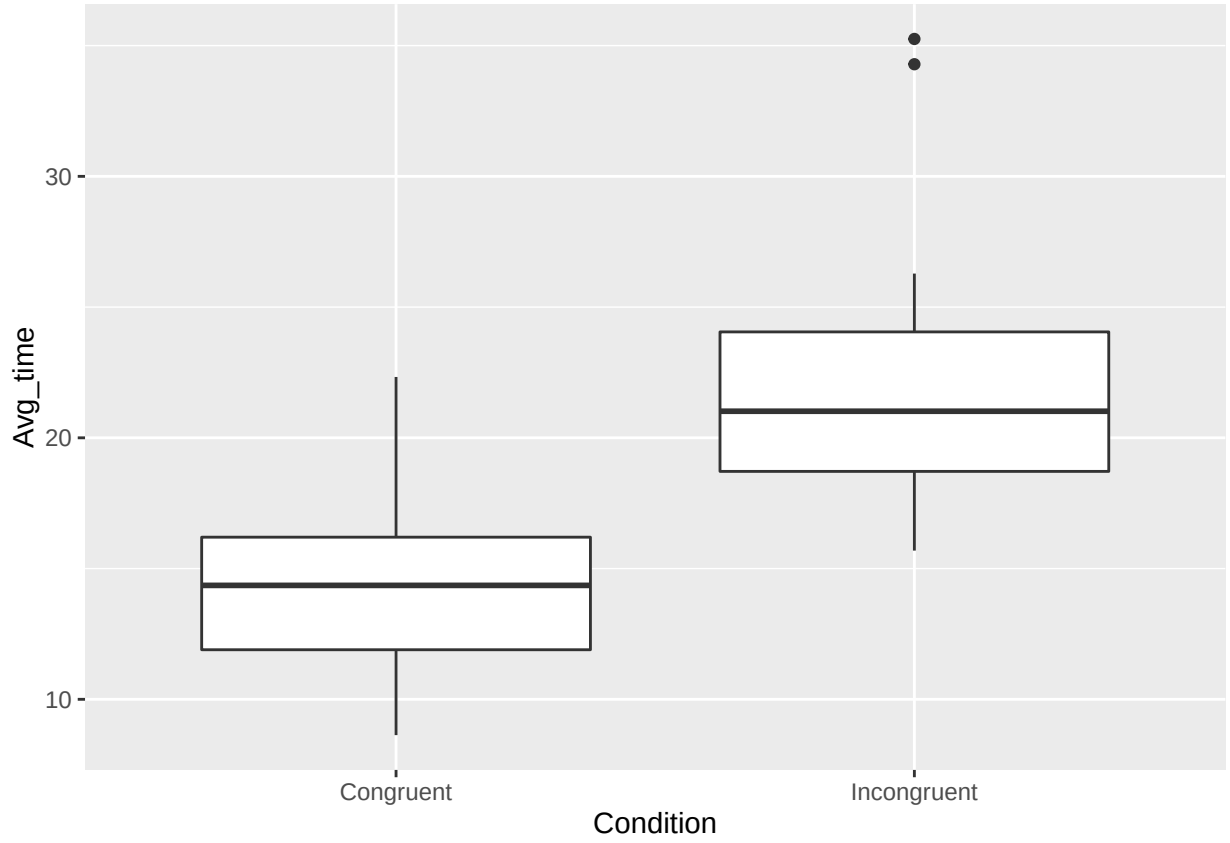
Histogram for incongruent condtion shows the presence of outliers. This indicates that, some participants took considerably more time than other participants in case of incongruent case. This might have pulled the mean time of response upwards.

The code below transforms the data. To get the sense of how transformed dataset looks, `head()` command is used. This is done to create parallel box plots based on two categories : Congruent and Incongruent.

```
#Transforming data to create Parallel Box Plots
Avg_time<-c(stroop$Congruent,stroop$Incongruent)
Condition<-c(rep("Congruent",nrow(stroop)),rep("Incongruent",nrow(stroop)))
stroop_trans<-data.frame(Avg_time,Condition)
stroop_trans$Condition<- factor(stroop_trans$Condition)
str(stroop_trans)
```

```
## 'data.frame':    48 obs. of  2 variables:
##  $ Avg_time : num  12.08 16.79 9.56 8.63 14.67 ...
##  $ Condition: Factor w/ 2 levels "Congruent","Incongruent": 1 1 1 1 1 1 1 1 1 1 ...
```

```
#Visualization of data using parallel box plots
ggplot(stroop_trans,aes(x=Condition,y=Avg_time))+geom_boxplot()
```



We can visually see that there is a clear difference between response time under both conditions. In fact, incongruent condition leads to significantly more time than congruent condition.

Outliers can be seen in case of incongruent condition.

Hypothesis Testing

We are performing the hypothesis test to see if the difference in response time under different condition is statistically significant. The obvious understanding and my personal experience with Stroop task gives me a sense that response time in case of incongruent test should always be more than congruent case.

Paired t-test has been used to perform the hypothesis. Both directional as well as non-directional tests are performed. It has been observed that data is not normally distributed. However, our dataset has 24 observations which is fairly big sample size. Thus, according to central limit theorem, sampling distribution will be nearly normal.

Null Hypothesis H_0:There is no significant difference between the time taken to respond in case of congruent words condition and incongruent words condition.

Alternative Hypothesis H_A: There is a significant difference between the time taken to respond in case of congruent words condition and incongruent words condition.

$$H_0 : \mu_c - \mu_I = 0$$
$$H_A : \mu_c - \mu_I \neq 0$$

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

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

Null Hypothesis H_0:There is no significant difference between the time taken to respond in case of congruent words condition and incongruent words condition.

Alternative Hypothesis H_A: The time taken to respond in case of congruent words condition is significantly less than the time take to respond in case of incongruent words condition.

$$H_0 : \mu_c - \mu_I = 0$$
$$H_A : \mu_c - \mu_I < 0$$

Here μ_c denotes sample mean of response time in case of congruent condition and μ_I denotes sample mean of response time in case of incongruent condition.

```
t.test(stroop$Congruent,stroop$Incongruent,paired=TRUE,alt="less")
```

```
##
##  Paired t-test
##
## data:  stroop$Congruent and stroop$Incongruent
## t = -8.0207, df = 23, p-value = 2.052e-08
```

```
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##       -Inf -6.262868
## sample estimates:
## mean of the differences
##               -7.964792
```

From the t test we performed above, we can conclude that the difference between response time under two conditions is statistically significant. Specifically, time taken to respond in case of congruent condition is significantly less than the time taken to respond in case of incongruent condition.

Further Investigation

The fundamental reason for Stroop Effect is `interference`. I feel that we are naturally quick at reading as compared to recognizing the color of the ink. So, if we are shown word BLUE in RED color, we will naturally and quickly read the text BLUE rather then identifying the color of the ink.

Another variation of stroop effect is emotional stroop.

In an emotional Stroop task, an individual is given negative emotional words like “grief,” “violence,” and “pain” mixed in with more neutral words like “clock,” “door,” and “shoe”.Just like in the original Stroop task, the words are colored and the individual is supposed to name the color. Research has revealed that individuals that are depressed are more likely to say the color of a negative word slower than the color of a neutral word. While both the emotional Stroop and the classic Stroop involve the need to suppress irrelevant or distracting information, there are differences between the two. The emotional Stroop effect emphasizes the conflict between the emotional relevance to the individual and the word; whereas, the classic Stroop effect examines the conflict between the incongruent color and word.