The Science of Decisions

Background

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.

Hypotheses

Independent variable: condition - congruent or incongruent

Dependent variable: time it takes to name the ink colors

H_0: There is no difference in the time it takes to name ink colors in the congruent list vs incongruent list, d = 0.

H_A: The time to name the ink colors for the incongruent list is greater than the congruent list, d > 0.

Load packages and data

```
library(ggplot2)
stroop <- read.csv("stroopdata.csv")
```

Descriptive Statistics

Sample Size

```
## [1] 24
```

mean time and standard deviation for congruent condition

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

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

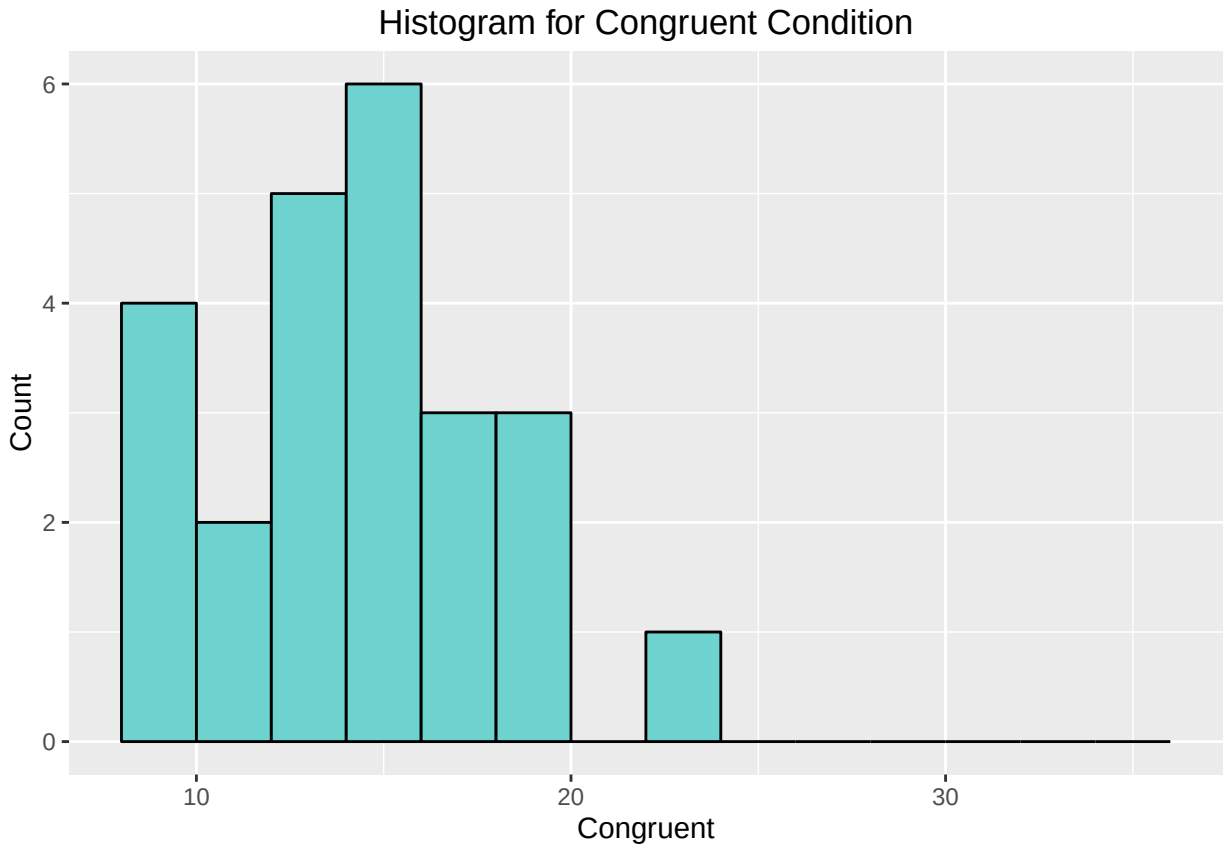
mean time and standard deviation for incongruent condition

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

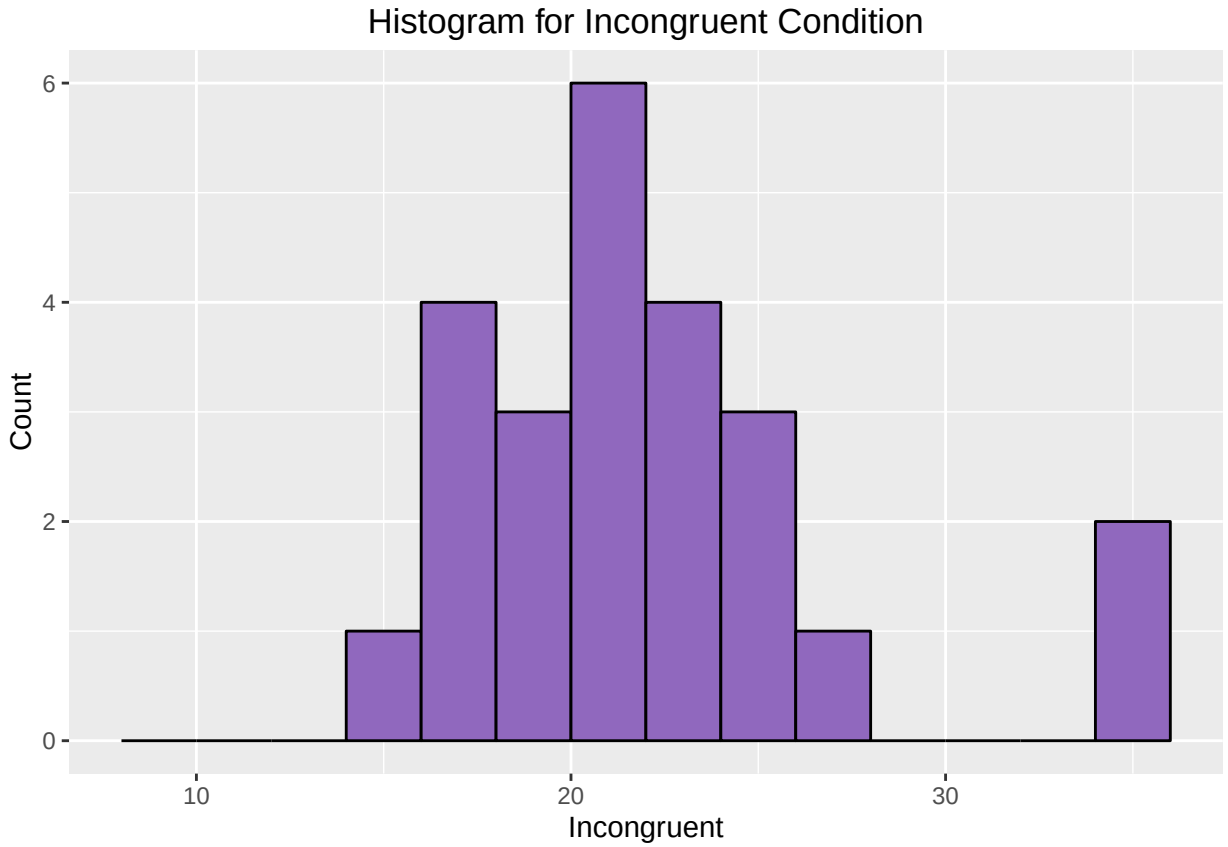
```
## [1] 4.797057
```

Plots to look at the distribution of the data

```
# congruent histogram
ggplot(data=stroop, aes(stroop$Congruent)) +
  geom_histogram(col="black",fill="#6ed3cf",breaks=seq(8, 36, by = 2)) +
  labs(title="Histogram for Congruent Condition") +
  labs(x="Congruent", y="Count")
```

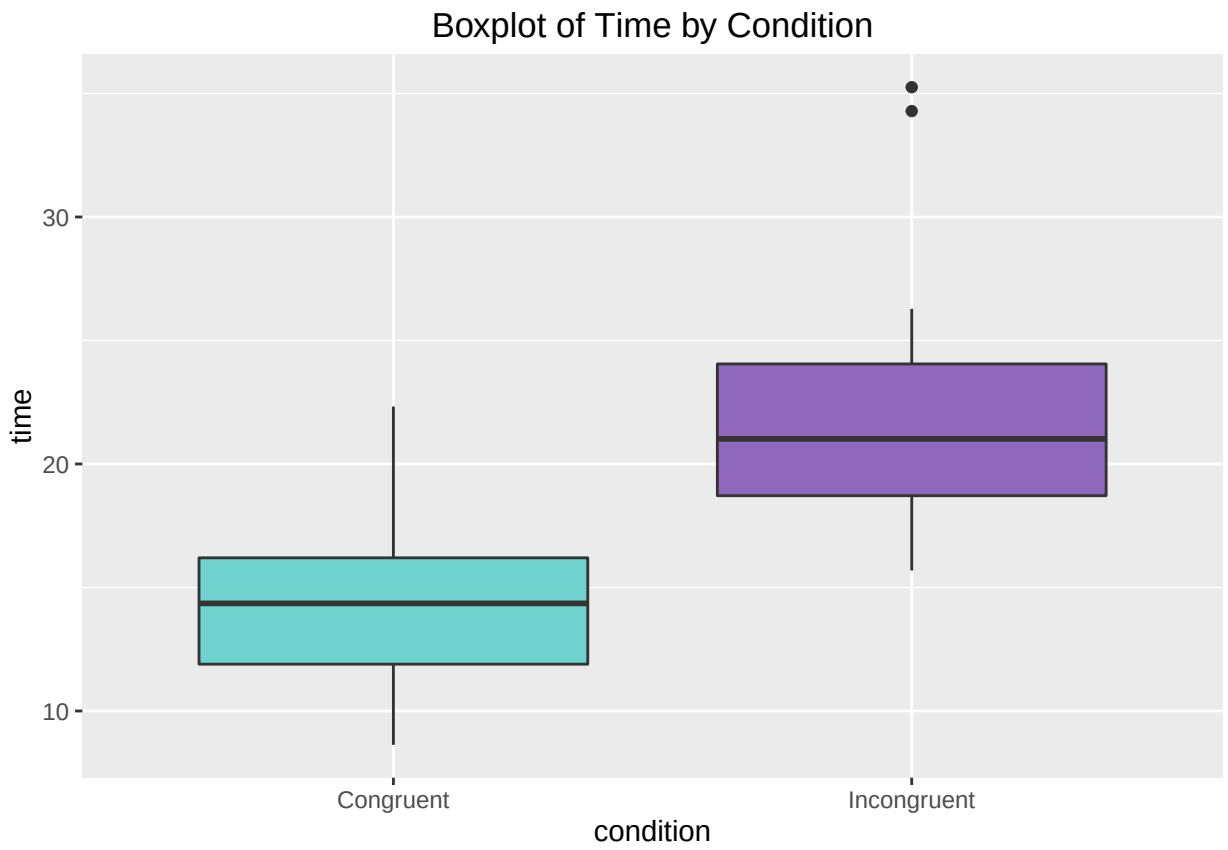


```r
# incongruent histogram
ggplot(data=stroop, aes(stroop$Incongruent)) +
  geom_histogram(col="black",fill="#9068be",breaks=seq(8, 36, by = 2)) +
  labs(title="Histogram for Incongruent Condition") +
  labs(x="Incongruent", y="Count")
```



```r
# transform data to produce boxplot
congruent <- data.frame(time=stroop$Congruent, condition="Congruent")
incongruent <- data.frame(time=stroop$Incongruent, condition="Incongruent")
stroop2 <- rbind(congruent,incongruent)
```

```r
ggplot(aes(x=condition, y=time), data=stroop2) +
  geom_boxplot(fill=c("#6ed3cf","#9068be")) +
  ggtitle("Boxplot of Time by Condition")
```



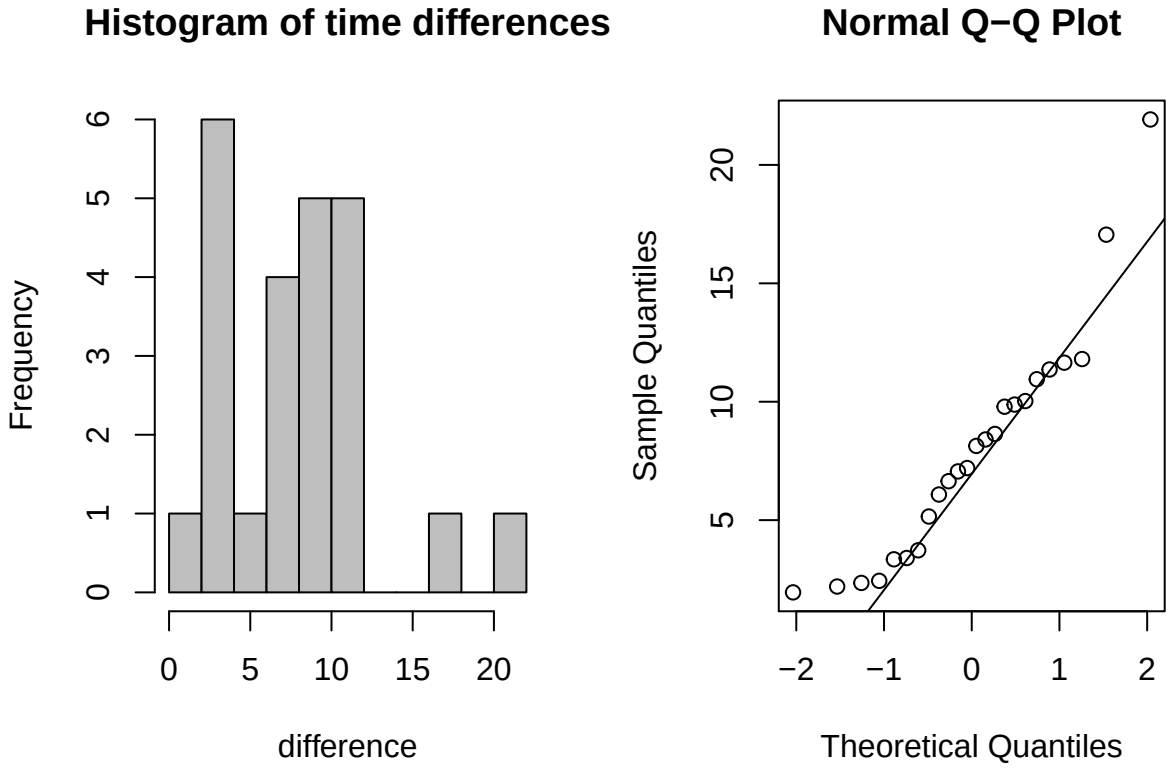
The congruent and incongruent histograms show that the data is approximately normally distributed. From the side-by-side box plot we can see that there appears to be a considerable difference in times between the two conditions. Both the incongruent histogram and boxplot show two potential outliers. However, due to the nature of the incongruent condition it seems likely that these are legitimate observations.

Perform Paired T-Test

A paired t-test was chosen to distinguish the proposed hypotheses because the data has matched pairs, meaning that a single subject performed both the congruent and incongruent tasks. Before performing the t-test we must check that the data meets the assumptions of a paired t-test.

The observations are independent, the sample is small, and the differences are nearly normal as seen in the plots below. The dots on the QQ plot almost line up along the line of equality, so this confirms that the distribution of the differences is nearly normal. Therefore, we can proceed with the paired t-test.

```
stroop$diff <- stroop$Incongruent - stroop$Congruent
par(mfrow=c(1,2))
hist(stroop$diff,breaks=12,col="gray",main="Histogram of time differences",
     xlab="difference")
qqnorm(stroop$diff)
qqline(stroop$diff)
```



First we'll calculate the t statistic by hand using this formula:

$$t = \frac{\bar{d}}{\sqrt{s^2/n}}$$

```
# mean difference
dbar <- mean(stroop$diff)
dbar
```

```
## [1] 7.964792
```

```
# standard error
se <- sd(stroop$diff) / sqrt(n)
se
```

```
## [1] 0.9930286
```

```
# calculate t
t <- dbar / se
t
```

```
## [1] 8.020707
```

```
# find p-value
1-pt(t,df=n-1)
```

```
## [1] 2.0515e-08
```

Our p-value is very close to zero and thus less than 0.05. We can reject the null hypothesis that there is no difference in the time it takes to name ink colors in the congruent list vs incongruent list.

Find t critical for a 95% confidence interval:

```
tcrit <- qt(0.975,df=n-1)
tcrit
```

```
## [1] 2.068658
```

```
lower <- dbar - tcrit * se
lower
```

```
## [1] 5.910555
```

```
upper <- dbar + tcrit * se
upper
```

```
## [1] 10.01903
```

Our confidence interval ranges from a lower bound of 5.91 to an upper bound of 10.02. Therefore, we are 95% confident that the true difference in congruent and incongruent times is between 5.91 and 10.02.

Lastly we will perform the paired t-test using R functions.

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

```
##
##  Paired t-test
##
## data:  stroop$Incongruent and stroop$Congruent
## t = 8.0207, df = 23, p-value = 2.052e-08
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  6.262868      Inf
## sample estimates:
## mean of the differences
##                7.964792
```

Remove alternative hypothesis option from the t.test command to get confidence interval.

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

```
##
##  Paired t-test
##
## data:  stroop$Incongruent and stroop$Congruent
## 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:
##   5.910555 10.019028
## sample estimates:
## mean of the differences
##                7.964792
```

Conclusion

In conclusion, the mean time to complete the task increased significantly for the incongruent condition (22.02 $\pm$ 4.78) vs the congruent condition (14.05 $\pm$ 3.56), t(23)=8.02, p $<$ 0.001, 95% CI [5.91, 10.02], d=7.96.