Class Activity #1

Sunny Lee

2021-01-22

Investigation: Do distracting Colors Influence the Time to complete a Game?

In this exercise, we would like to explore Two-Sample t-Test to compare the completion time of two independent populations. For this exercise, we will be using the Game data to understand if distracting colors affect the completion time of a Game.

1) The researchers hoped to determine if distracting colors influenced college students response times when playing a computerized game. Write out in words and symbols appropriate null and alternative hypothesis.

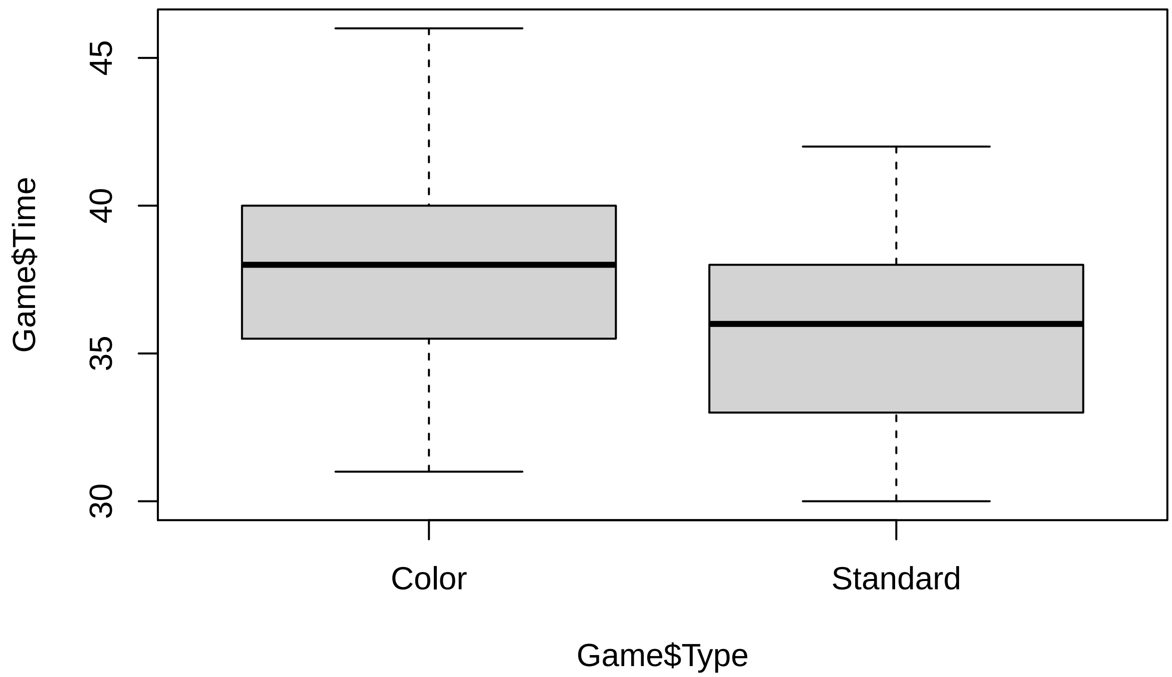
Solution

Let μ_1 be the completion time for a standard game and μ_2 be the comlpetion time for a color game. $H_o : \mu_1 = \mu_2$

$H_a : \mu_1 \neq \mu_2$

2) Create an indivial value plot or a box plot of the Game data from this study. Describe the graph. For example, does it look as if the groups have equal means or standard deviations? Are there any unusual observations in the data set? Calculate the mean and standard deviation of the colored and standard data.

```
Game <- read.csv("C2 Games1.csv")
#Independent, yes, we dont see any students in both game types
boxplot(Game$Time~Game$Type)
```



From the box plot, we see that the two groups have different means and seem to have different standard deviations. We can also see this by finding the actual mean and standard deviation in R:

```
Standard <- subset(Game, Game$Type == "Standard")
Color <- subset(Game, Game$Type == "Color")

mean(Standard$Time)
```

```
## [1] 35.55
```

```
sd(Standard$Time)
```

```
## [1] 3.394655
```

```
mean(Color$Time)
```

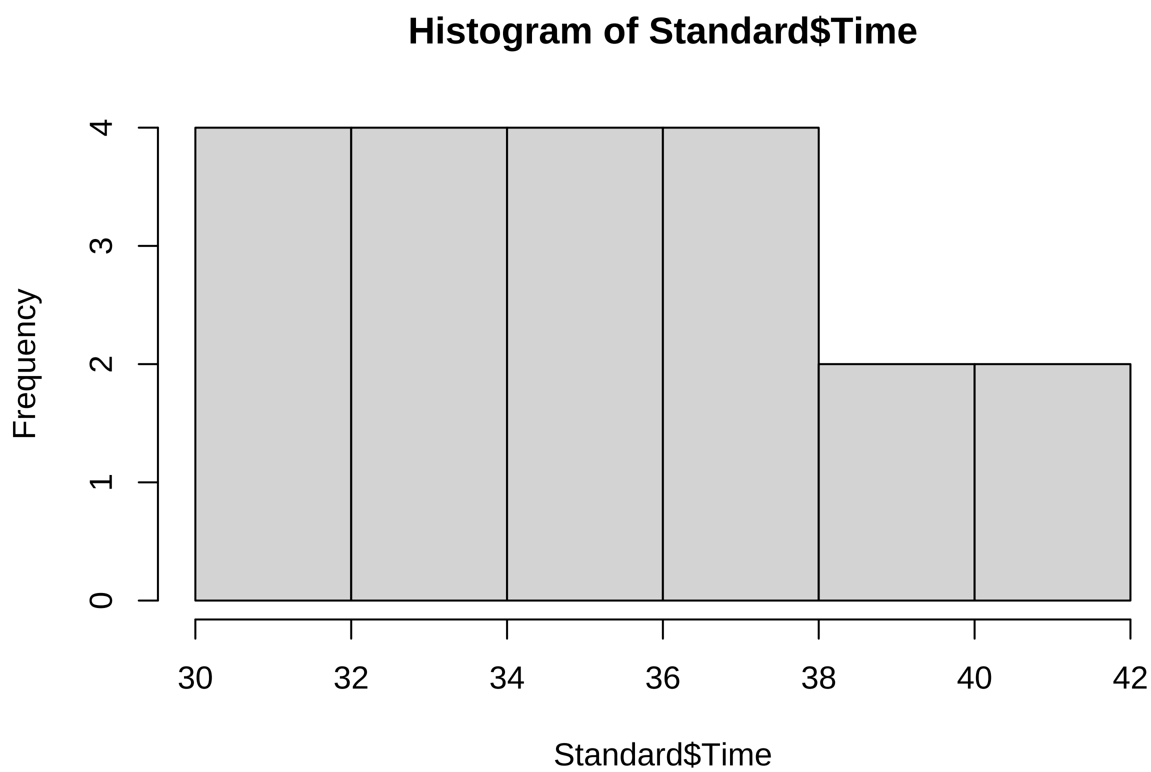
```
## [1] 38.1
```

```
sd(Color$Time)
```

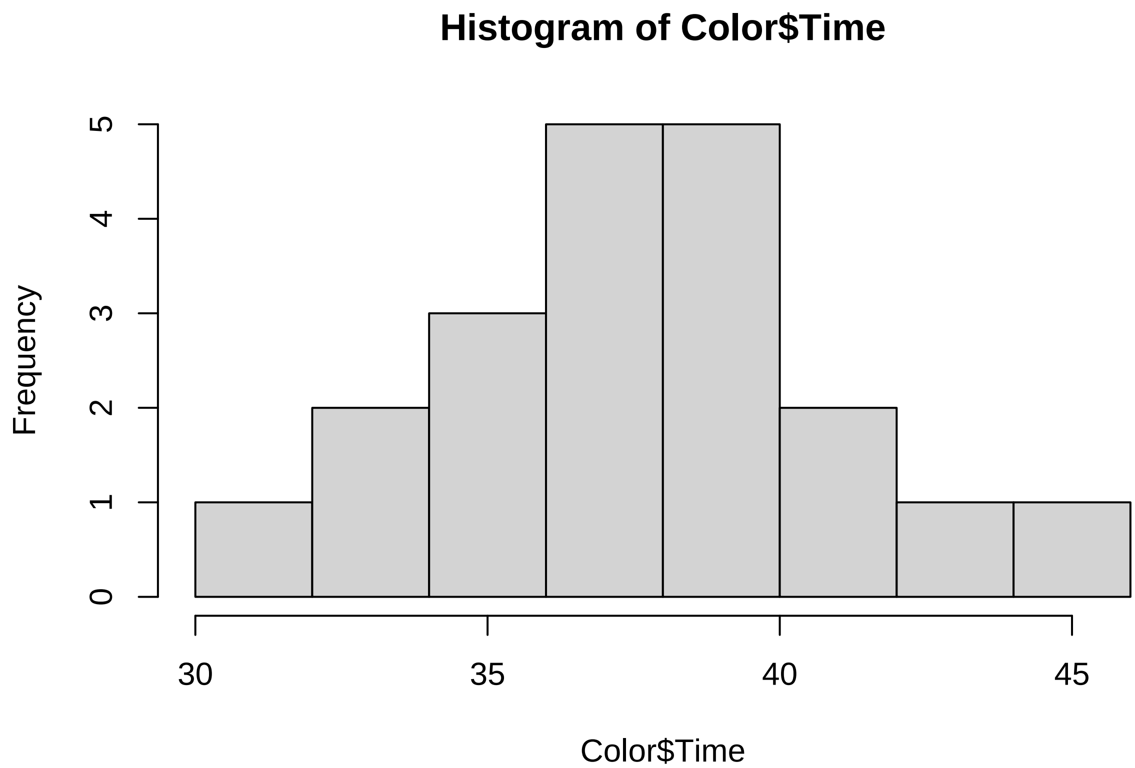
```
## [1] 3.654845
```

From the histogram of the Standard and Color group times, we also see only the Color group looks normal whereas the Standard time histogram does not.

```
hist(Standard$Time)
```

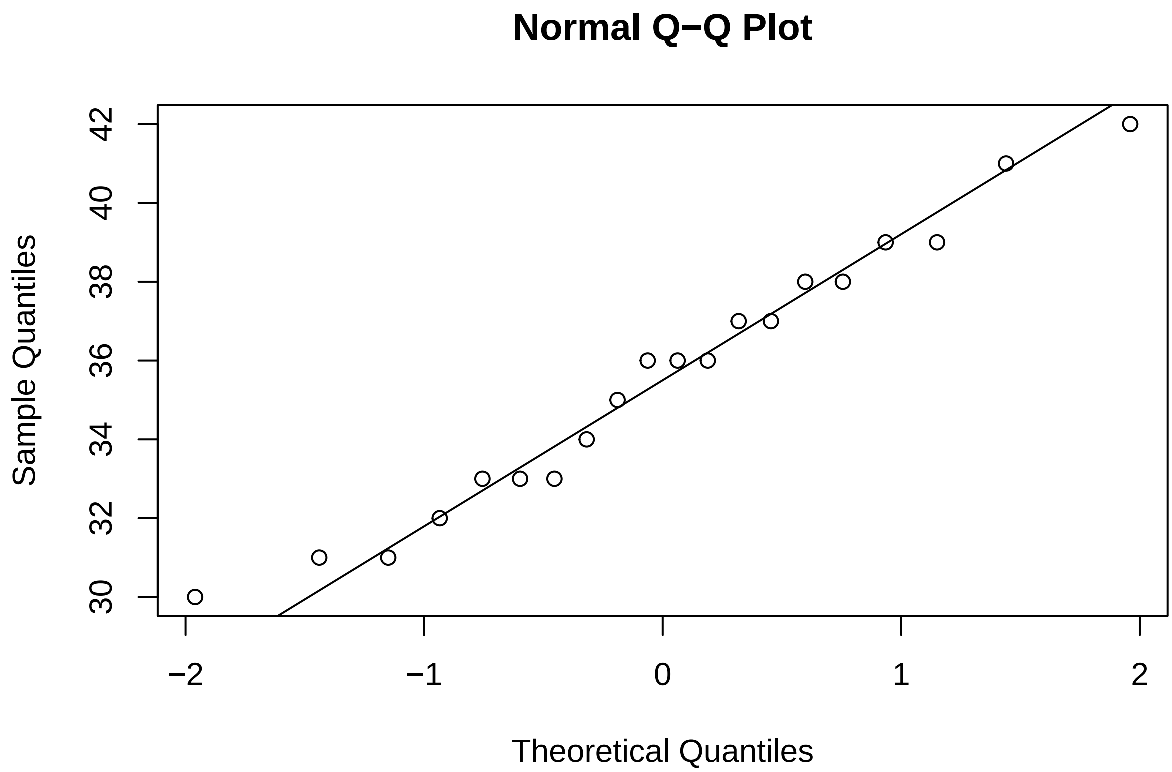



```
hist(Color$Time)
```

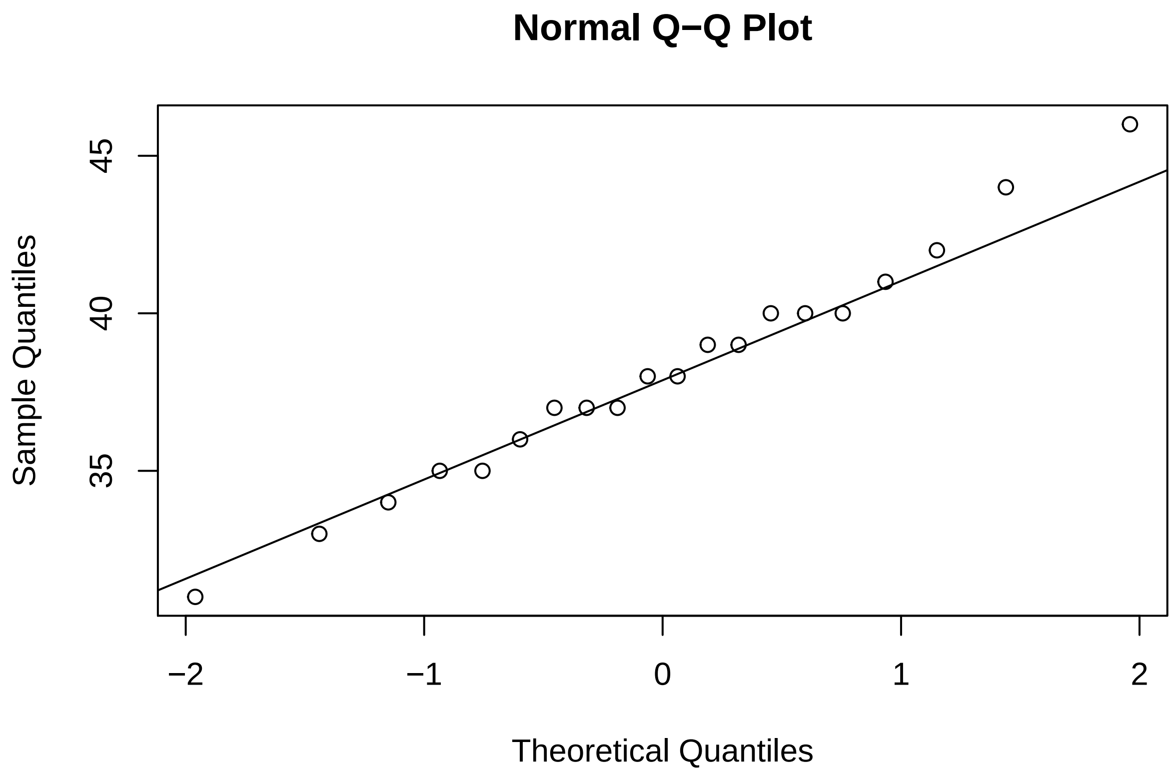


For outliers, we can use the qqnorm graphs for each Standard and Color times. We see that most of the points on both Standard and Color stay close to the line, with only one or two outliers.

```
qqnorm(Standard$Time)
qqline(Standard$Time)
```




```
qqnorm(Color$Time)
qqline(Color$Time)
```



3) Calculate the residuals in the Games data. Plot a histogram of the residuals (or create a normal probability plot of the residuals). Do the residuals appear to be somewhat normally distributed?

To calculate the residuals, we first take the residuals of the Standard and Color time:

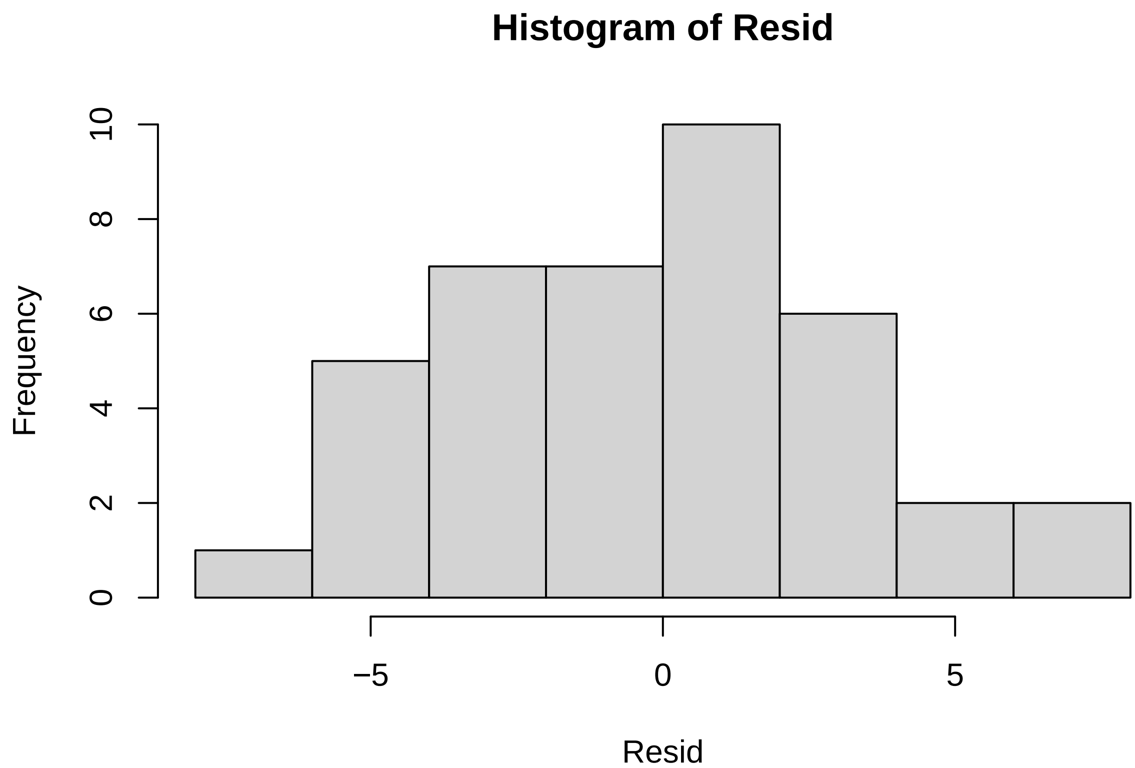
```
Standard_Time <- subset(Game, Game$Type == "Standard")$Time
Color_Time <- subset(Game, Game$Type == "Color")$Time

Resid_Standard <- Standard_Time-mean(Standard_Time)
Resid_Color <- Color_Time-mean(Color_Time)
```

and then combine:

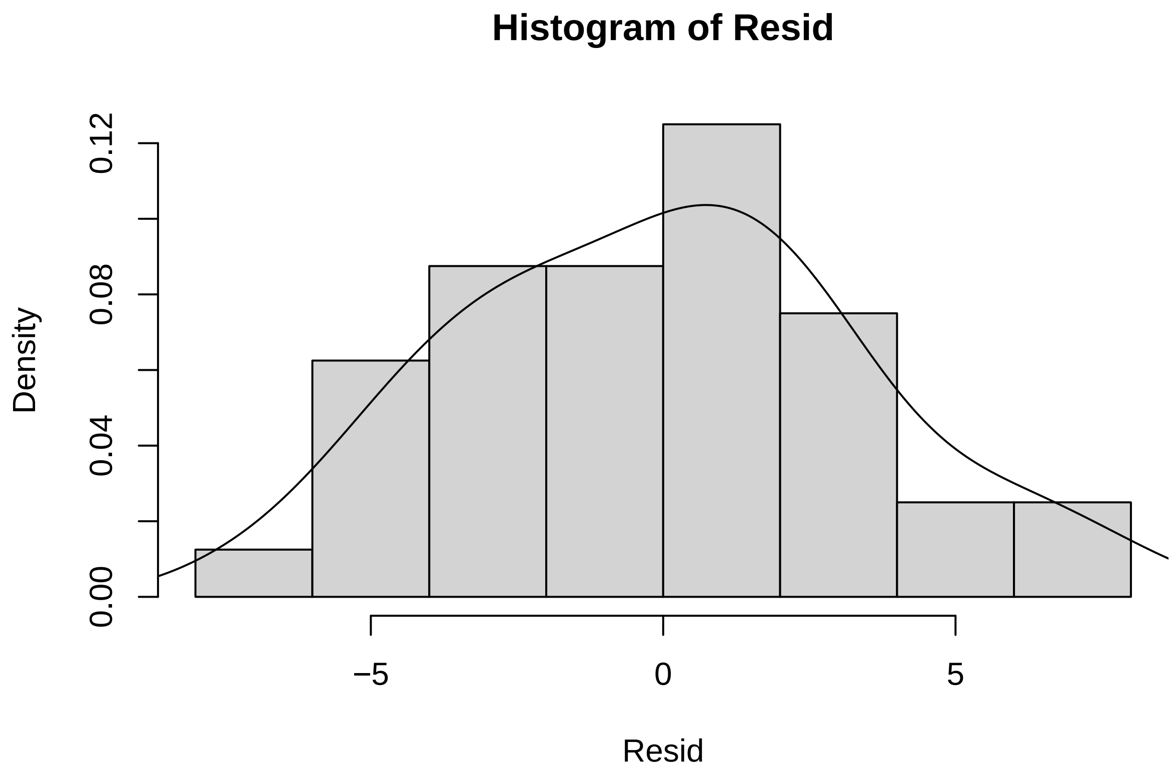
```
Resid <- c(Resid_Standard, Resid_Color)
```

When we plot a histogram of the residuals, we find that the residuals look somewhat normally distributed.

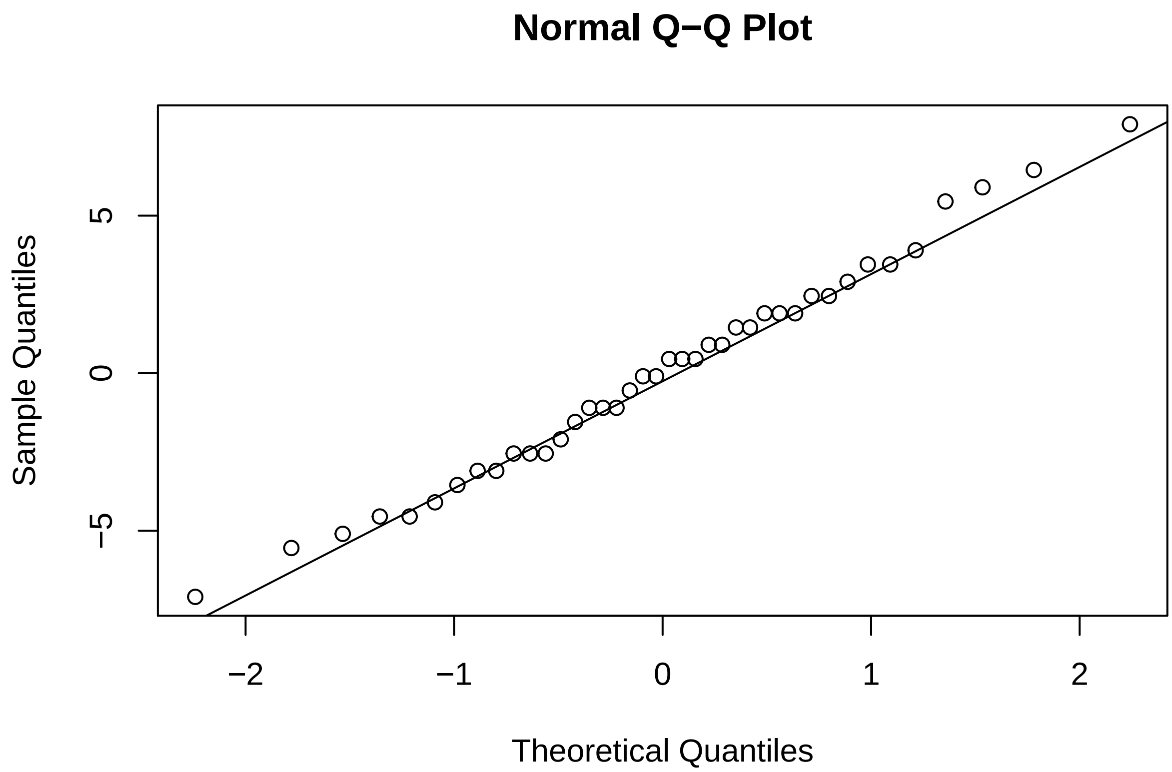
```
hist(Resid)
```




```
hist(Resid, freq=FALSE)
points(density(Resid), type="l")
```




```
qqnorm(Resid)
qqline(Resid)
```



4) Use the informal test to determine if the equal variance assumption is appropriate for this study.

```
max(var(Standard$Time), var(Color$Time)) / min(var(Standard$Time), var(Color$Time))
```

```
## [1] 1.159169
```

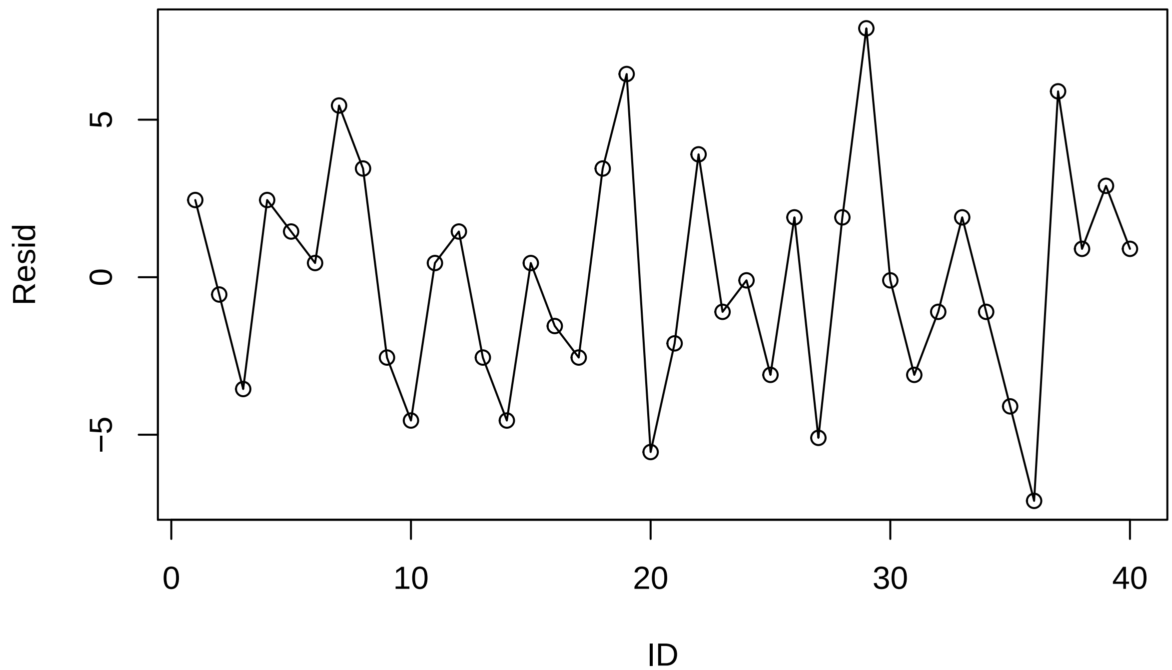
```
max(sd(Standard$Time), sd(Color$Time)) / min(sd(Standard$Time), sd(Color$Time))
```

```
## [1] 1.076647
```

Since the max standard deviation/variance over the min standard deviation/variance is less than 2, informally, we can say the variance is equal.

5) The variable studentID represents the order in which the games were played. Plot the residuals versus the order of the data to determine if any patterns exist that may indicate that the observations are not independent.

```
ID <- Game$studentID

plot(Resid~ID)
points(Resid~ID, type = "l")
```



Here we find there is no pattern to the residuals and the ID.

6) Use statistical software to conduct a two- sample t-test (assuming equal variance) and find the p-value corresponding to this statistic. In addition, use software to calculate a 95% confidence interval for the difference between thet two means ($\mu_1 - \mu_2$).

```
#t.test(Game$Time~Game$Type, var.equal=TRUE, paired=FALSE, conf.level=.95, mu=0, alternative = "two.sid

x <- rnorm(50)
y <- rnorm(50)

t.test(x, y, paired = FALSE, var.equal = TRUE, conf.level = .95, mu = 0, alternative = "two.sided")
```

```
##
##  Two Sample t-test
##
## data:  x and y
## t = 1.41, df = 98, p-value = 0.1617
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1141934  0.6747739
## sample estimates:
##   mean of x   mean of y
##  0.18685199 -0.09343826
```