Class Activity #2

Write Your Name

2021-01-27

Section 2.3: Linear Regression and Hypothesis Test for the Game data

1) Use the software instructions and the Game data to create indicator variables where $x = 1$ represents the color distracter game and $x = 0$ represents the standard game. Develop a regression model using Time as the response and the indicator variable as the explanatory variable.

```
Game <- read.csv("C2 Games1.csv")
```

Since Standard and Color are not numbers, we turn the Type column into a boolean array. We then multiply that array by one in order to turn the trues into 1 and false into 0.

```
D <- (Game$Type == "Color")*1 # boolean array for color = 1, standard = 0
```

Now that we have numerical data to work with, we can fit a linear model (lm) to our data and calling the summary function on our linear model Fit, we see very useful information about Residuals and coefficients.

```
Fit <- lm(Game$Time~D) # least square linear regression

summary(Fit)
```

```
##
## Call:
## lm(formula = Game$Time ~ D)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -7.100 -2.550  0.175  2.038  7.900
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  35.5500     0.7887  45.074   <2e-16 ***
## D             2.5500     1.1154   2.286   0.0279 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.527 on 38 degrees of freedom
## Multiple R-squared:  0.1209, Adjusted R-squared:  0.09778
## F-statistic: 5.227 on 1 and 38 DF,  p-value: 0.02791
```

2) Use statistical software to calculate the t-statistic and p-value for the hypothesis tests. $H_o : \beta_1 = 0$ versus $H_a : \beta_1 \neq 0$. In addition, construct a 95% confidence interval for β_1. Based on these statistics, can you conclude that the coefficient, β_1, is significantly different from zero?

From the summary above, we find that the Pr(>|t|) for β_1 is less than our $\alpha = .05$, thus, we must reject the null hypothesis. In order to construct a 95% confidence interval, we must take the estimate for β_1 and add/subtract the $t_{df,\frac{\alpha}{2}} sd(\hat{\beta})$ which can also obtain from the summary above.

```
Lower <- 2.55 + qt(.025, df = 38) * 1.1154

Upper <- 2.55 - qt(.025, df = 38) * 1.1154

cat(sprintf("Lower: %.8f\n", Lower))
```

```
## Lower: 0.29199075
```

```
cat(sprintf("Upper: %.8f", Upper))
```

```
## Upper: 4.80800925
```

3) Repeat the two previous questions, but use an indicator where $x = 1$ represents the standard game and $x = 0$ represents the color distracter game. Compare the regression line, hypothesis test, and p-value to those from the previous questions.

Repeating the steps from above:

```
D1 <- (Game$Type == "Standard")*1

Fit1 <- lm(Game$Time~D1) # least square linear regression

summary(Fit1)
```

```
##
## Call:
## lm(formula = Game$Time ~ D1)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -7.100 -2.550  0.175  2.038  7.900
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  38.1000     0.7887  48.308   <2e-16 ***
## D1           -2.5500     1.1154  -2.286   0.0279 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.527 on 38 degrees of freedom
## Multiple R-squared:  0.1209, Adjusted R-squared:  0.09778
## F-statistic: 5.227 on 1 and 38 DF,  p-value: 0.02791
```

```
Lower <- -2.55 + qt(.025, df = 38) * 1.1154

Upper <- -2.55 - qt(.025, df = 38) * 1.1154

cat(sprintf("Lower: %.8f\n", Lower))
```

```
## Lower: -4.80800925
```

```
cat(sprintf("Upper: %.8f", Upper))
```

```
## Upper: -0.29199075
```

Here we see that everything is the same except that now our interval is negative.

4) Calculate the residuals from the regression line in question 1. Plot a histogram of the residual (or create a normal probability plot of the residuals). In addition, create a residual versus order plot and use the

informal test to determine if the equal variance assumption is appropriate for this study. Compare these plots to the residual plots created for the two sample t-test.

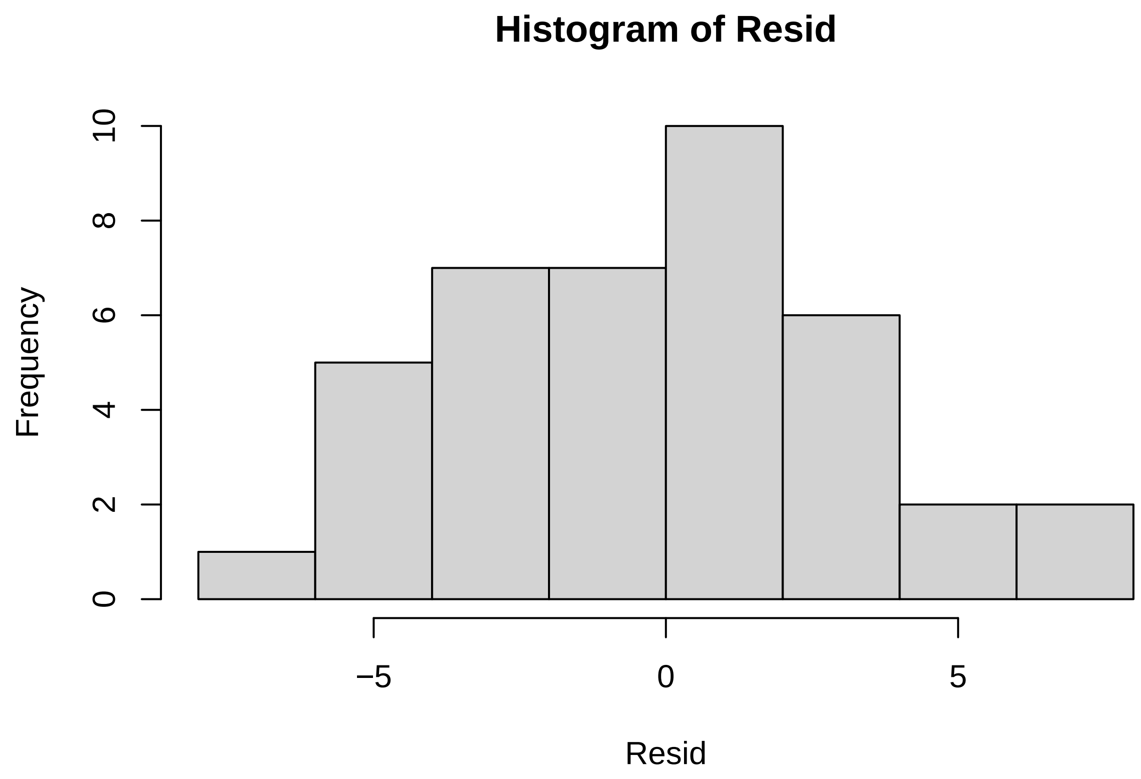
Earlier, we fit a linear model onto our data using the lm() function and called it Fit. Looking at the summary of Fit, we find that Fit has many variables which are very useful to us, one of which is the residuals. Therefore, since Fit contains information about our residuals we can simply call it from our Fit variable.

```
#using variables from questions 1 and 2 not from 3
Resid <- Fit$residuals
Resid
```

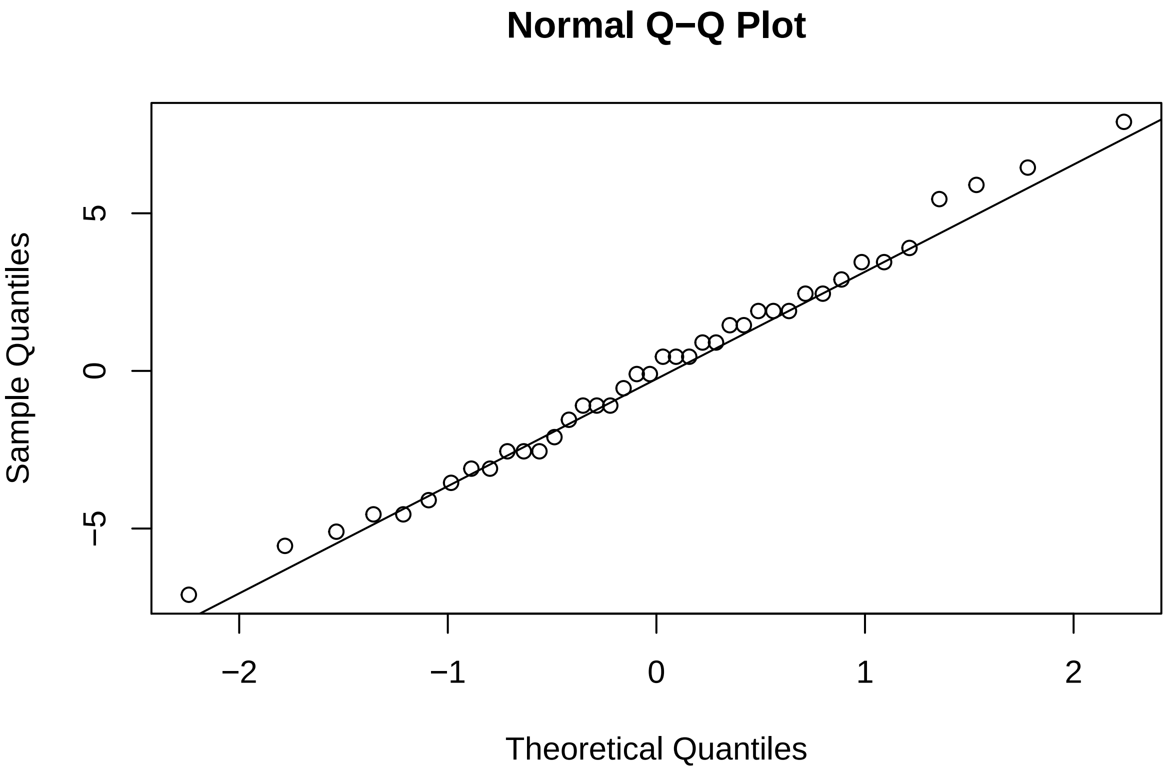
```
##     1     2     3     4     5     6     7     8     9    10    11    12    13 
##  2.45 -2.10  3.90 -0.55 -3.55 -1.10 -0.10  2.45  1.45  0.45 -3.10  1.90  5.45 
##    14    15    16    17    18    19    20    21    22    23    24    25    26 
##  3.45 -5.10  1.90  7.90 -2.55 -0.10 -4.55  0.45  1.45 -3.10 -2.55 -1.10  1.90 
##    27    28    29    30    31    32    33    34    35    36    37    38    39 
## -1.10 -4.55  0.45 -1.55 -4.10 -2.55 -7.10  5.90  0.90  3.45  6.45  2.90  0.90 
##    40 
## -5.55
```

Histogram of the residuals:

```
hist(Resid)
```



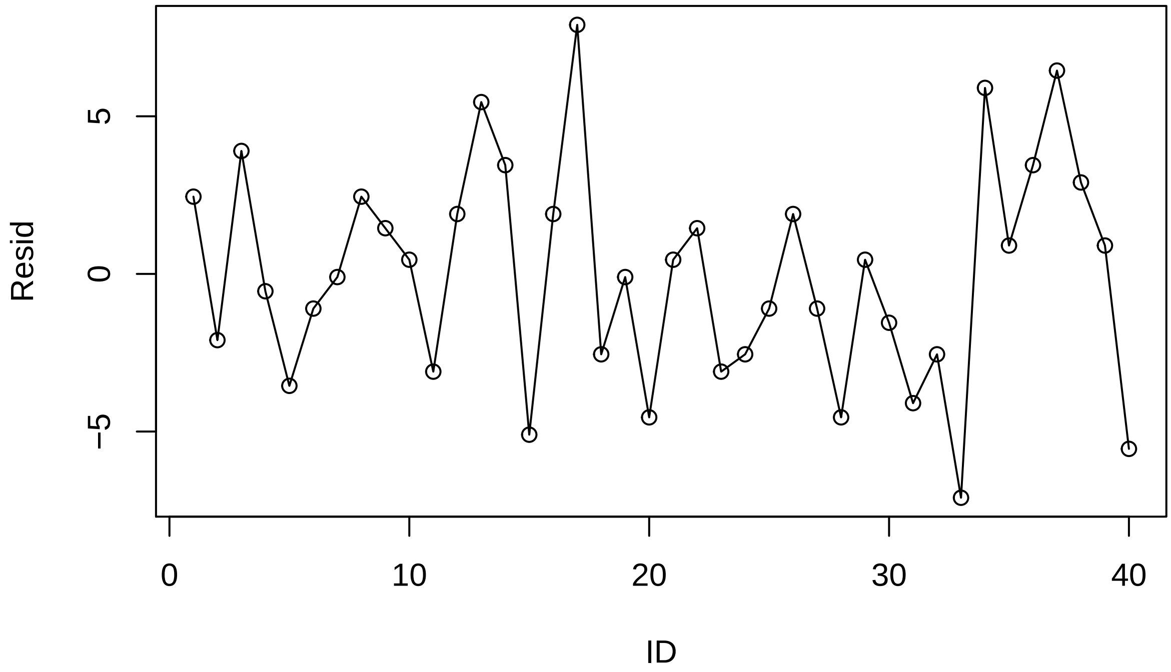
```
qqnorm(Resid)
qqline(Resid)
```



A plot of the residuals according to the studentID, to make sure that there is no pattern which exists in our residuals.

```
ID <- Game$studentID

plot(Resid~ID)
points(Resid~ID, type = "l") #you can use any order
```



```
mean(Resid)
```

```
## [1] -1.67184e-17
```

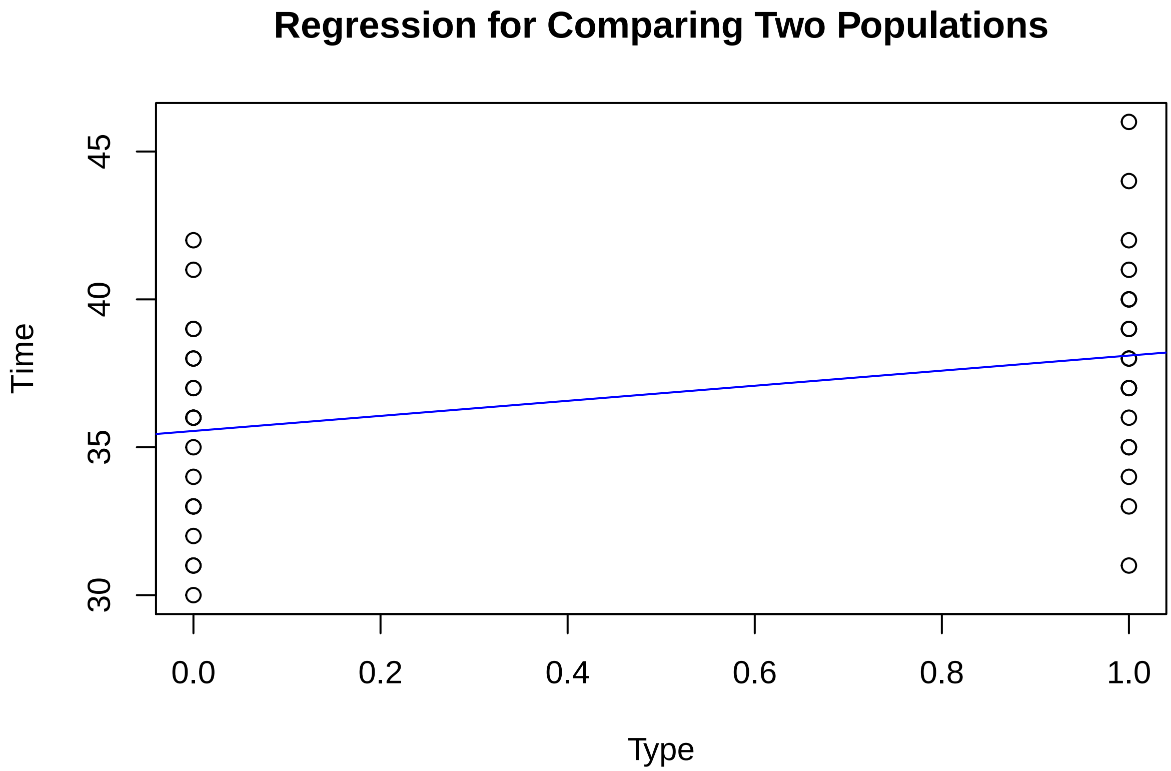
Using the informal test to check variance equivalence, we find that since our max var over our min var is less than 4, we assume our variance in the two groups are equal:

```
Standard <- subset(Game, Game$Type == "Standard")
Color <- subset(Game, Game$Type == "Color")
max(var(Standard$Time), var(Color$Time)) / min(var(Standard$Time), var(Color$Time)) < 4
```

```
## [1] TRUE
```

5) Create a scatterplot with the regression line in question 1. Use the graph to give an interpretation of the slope and y-intercept, b_1 and b_0 in the context of the game study.

```
plot(Game$Time~D, ylab="Time", xlab="Type", main="Regression for Comparing Two Populations")
abline(Fit, col = "Blue")
```



β_1 is the rate of change of y for every $D_i = 1$ and β_0 is the average of y when $D_i = 0$