Class Activity #3

Sunny Lee

2021-02-03

Section 2.4: Analysis of variance to compare Population Means for the Game data.

1) Use the game data set to calclulate $\bar{y}_{..}$, $\bar{y}_{1.}$, and $\bar{y}_{2.}$. Our grand mean we can simply find by taking the mean of all the time in the dataset. Our Standard and color means we can find by splitting our data into color and standard games and then finding the mean for each.

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

grand_mean <- mean(game$Time)

standard_game <- mean(game[game$Type == "Standard", ]$Time)

color_game <- mean(game[game$Type == "Color", ]$Time)
```

2) Estimate the effect sizes for the color distracter game and the standard game. To find the effect sizes, we simply take the means of standard and color games and subtract the grand mean.

```
std_effect <- standard_game - grand_mean

color_effect <- color_game - grand_mean
```

3) Calculate the residuals for the 20^{th} observation from the standard group, $\hat{\epsilon}_{2,20}$

To find the residual for the 20^{th} observation from the standard group, we can take the 20^{th} observation of the standard group and subtract the mean time of the standard group. We can also find the residuals of every observation in the color and standard groups:

```
standard <- game[game$Type == "Standard", ]

error1 <- standard[, 3]-standard_game

cat(sprintf("The error for the 20th observation in the standard group is: %.2f\n", error1[20]))
```

```
## The error for the 20th observation in the standard group is: -5.55
```

```
color <- game[game$Type == "Color", ]
error2 <- color[, 3]-color_game

error <- c(error1, error2)

error
```

```
##  [1]  2.45 -0.55 -3.55  2.45  1.45  0.45  5.45  3.45 -2.55 -4.55  0.45  1.45
## [13] -2.55 -4.55  0.45 -1.55 -2.55  3.45  6.45 -5.55 -2.10  3.90 -1.10 -0.10
## [25] -3.10  1.90 -5.10  1.90  7.90 -0.10 -3.10 -1.10  1.90 -1.10 -4.10 -7.10
## [37]  5.90  0.90  2.90  0.90
```

```r
SSE <- sum(error^2) #sum of square error

#sum of square total is the square (yij - grand_mean)
total <- game[, 3]-grand_mean
SST <- sum(total^2)

SSGroup <- SST-SSE
SSGroup
```

```
## [1] 65.025
```

4) Use statistical software to calculate the F-statistic and the p-value. Use the p-value to draw conclusions from the study. In order to calculate the F-statistic and the p-value, we can use use the aov() function which will make a ANOVA chart for us with the data we give it:

```r
#aov() is analysis of variance for comparing means
model <- aov(game$Time~game$Type)

summary(model)
```

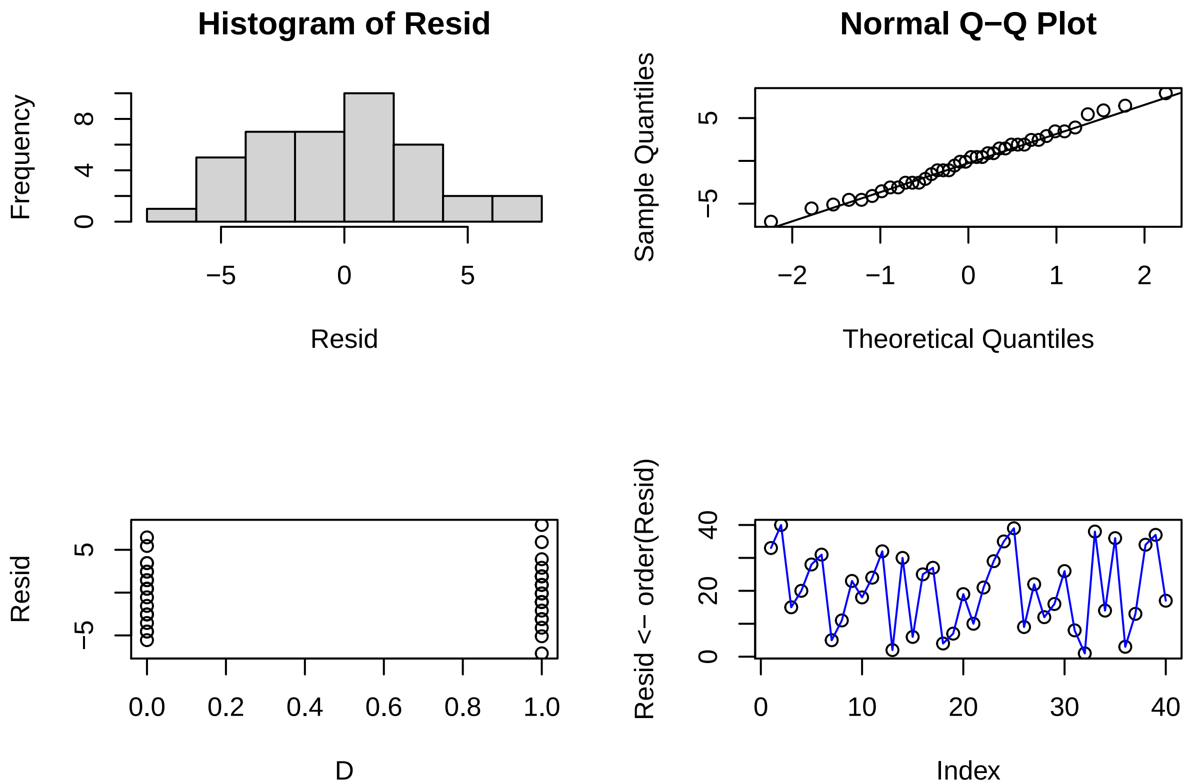
```
##             Df Sum Sq Mean Sq F value Pr(>F)
## game$Type    1   65.0   65.02   5.227 0.0279 *
## Residuals   38  472.8   12.44
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

And we see that our p-value is exactly the same as the p-value as the two sample t-test and linear regression.

5) How does the p-value in the ANOVA compare to the p-value you found for the two-sample t-test and the regression model? They are all the same, with no difference.

6) Check model assumptions by creating a histogram of the residuals, a plot of the residuals versus the type of game, and a plot of the residuals versus the order. Are the residuals approximately normal? are the residuals variances similar for the two factor levels? Are there patterns in the residual plots that might indicate that the residuals are not iid?

```r
Resid <- model$residuals
par(mfrow = c(2, 2))
hist(Resid)
qqnorm(Resid)
qqline(Resid)
D <- (game$Type == "Color")*1
plot(Resid~D)
plot(Resid <- order(Resid))
points(Resid, col = "blue", type = "l")
```



From the histogram and the qqnorm functions, we see the data seems to be quite close to being normally distributed. From the plot of the residuals versus the type of game, we find the variance seems quite similar, however the Color game type seems to have a bit more variance. From the residuals versus the order plot, we find the residuals have no pattern to them, thus we can conclude that the residuals are iid.