Yan Zhang's Anova Moon Phase Project

Yan Zhang

November 3, 2015

In this project, I am going to use the data MentalHealth from Stat2Data to check out if phases of moon will affect hospital admissions. This project in cludes three tests, which are one One-way anova test and two Two-way anova test.

One way anova:

First of all, I need to load the MentalHealth data from Stat2Data library:

```
library(Stat2Data)
data(MentalHealth)
head(MentalHealth)

##   Month   Moon Admission
## 1   Aug Before       6.4
## 2   Sep Before       7.1
## 3   Oct Before       6.5
## 4   Nov Before       8.6
## 5   Dec Before       8.1
## 6   Jan Before      10.4

levels(MentalHealth$Moon)

## [1] "After"  "Before" "During"
```

In this step, I am going to creat a new data that only contains months I need:

```
###subset corresponding monthes
Winter = c("Dec", "Jan", "Feb")
Fall = c("Sep", "Oct", "Nov")
sub.Month=as.factor(c(Winter, Fall))
lab4.data=subset(MentalHealth, Month %in% sub.Month)

head(lab4.data)

##   Month   Moon Admission
## 2   Sep Before       7.1
## 3   Oct Before       6.5
## 4   Nov Before       8.6
## 5   Dec Before       8.1
## 6   Jan Before      10.4
## 7   Feb Before      11.5
```

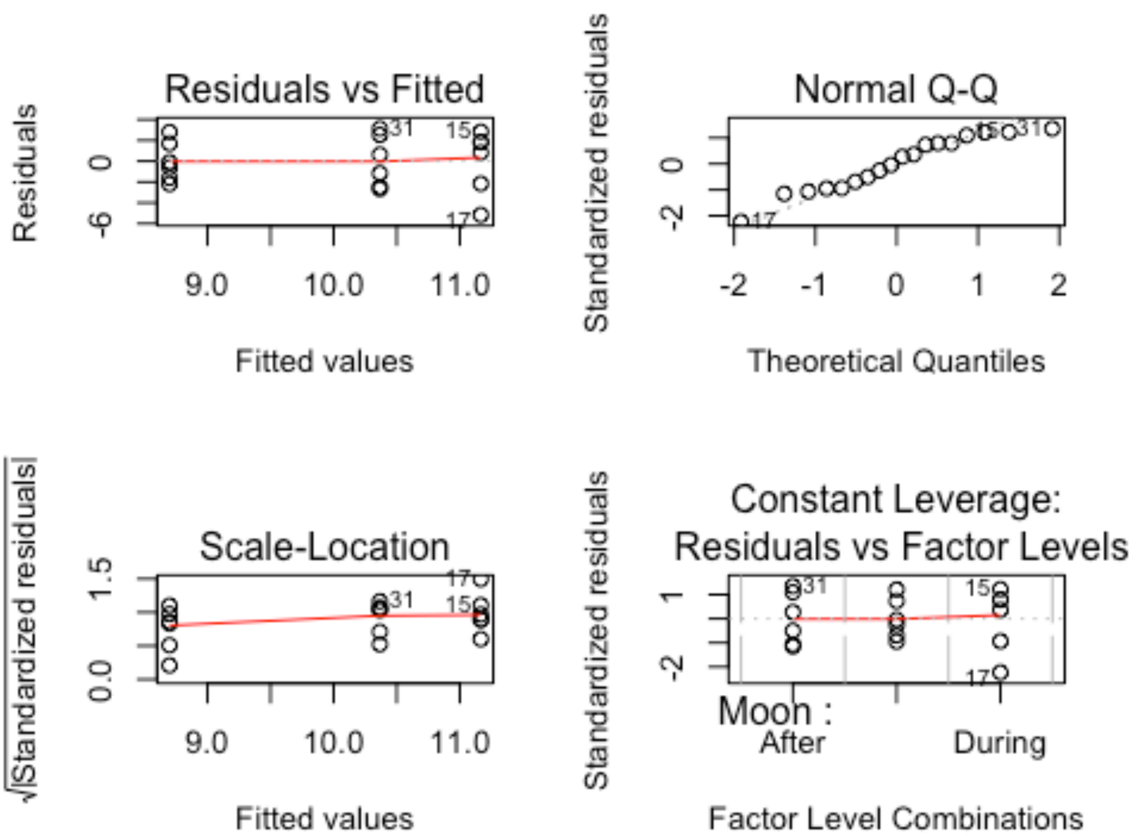
Since there is only one category, I am going to use one way anova, but I need to check assumption first:

Here is my hypotheses:

H_0: $\mu_1 = \mu_2$ H_a: At least one μ differ from the rest

```
foo1<-aov(Admission~Moon,data=lab4.data)

par(mfrow=c(2,2))
plot(foo1)
```




```
par(mfrow=c(1,1))
```

From the plots above, we can see the model has constant variance from Residuals vs Fitted plot, but the data isn't perfectly normal but it's still acceptable in QQ plot and we assume the data is randomlly collected, so the assumption is met.

From the pervious test, I will do the on way anova test to test out if the Moon variable is significant:

```
anova(foo1)
```

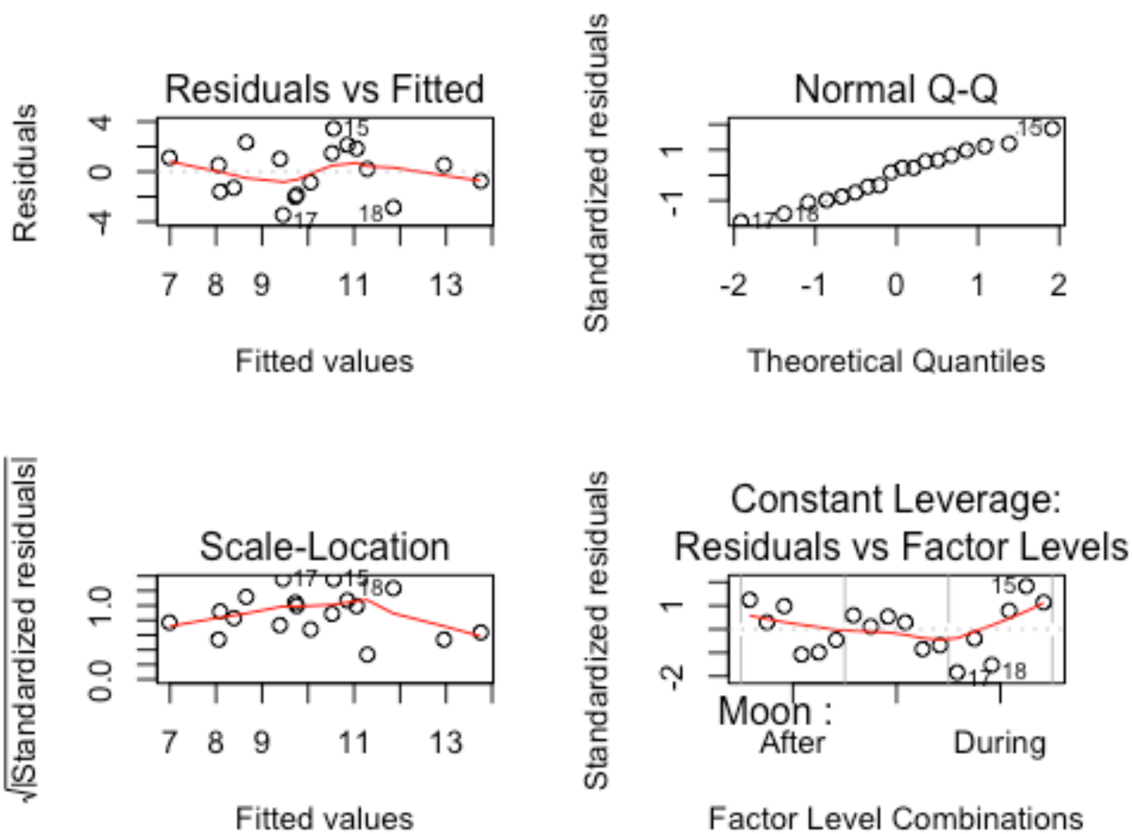
```
## Analysis of Variance Table
##
## Response: Admission
##           Df Sum Sq Mean Sq F value Pr(>F)
## Moon       2 19.004  9.5022  1.4766 0.2598
## Residuals 15 96.527  6.4351
```

From the anova test above, we can see the p value is not significant so we can say Moon phase doesn't affect hospital admissions.

Two way anova:

```
foo2<-aov(Admission~Moon+Month,data = lab4.data)

par(mfrow=c(2,2))
plot(foo2)
```




```
par(mfrow=c(1,1))
```

Here is my hypotheses:

$H_0: \mu_1 = \mu_2 = \mu_3$ H_a: At least one μ differ from the rest

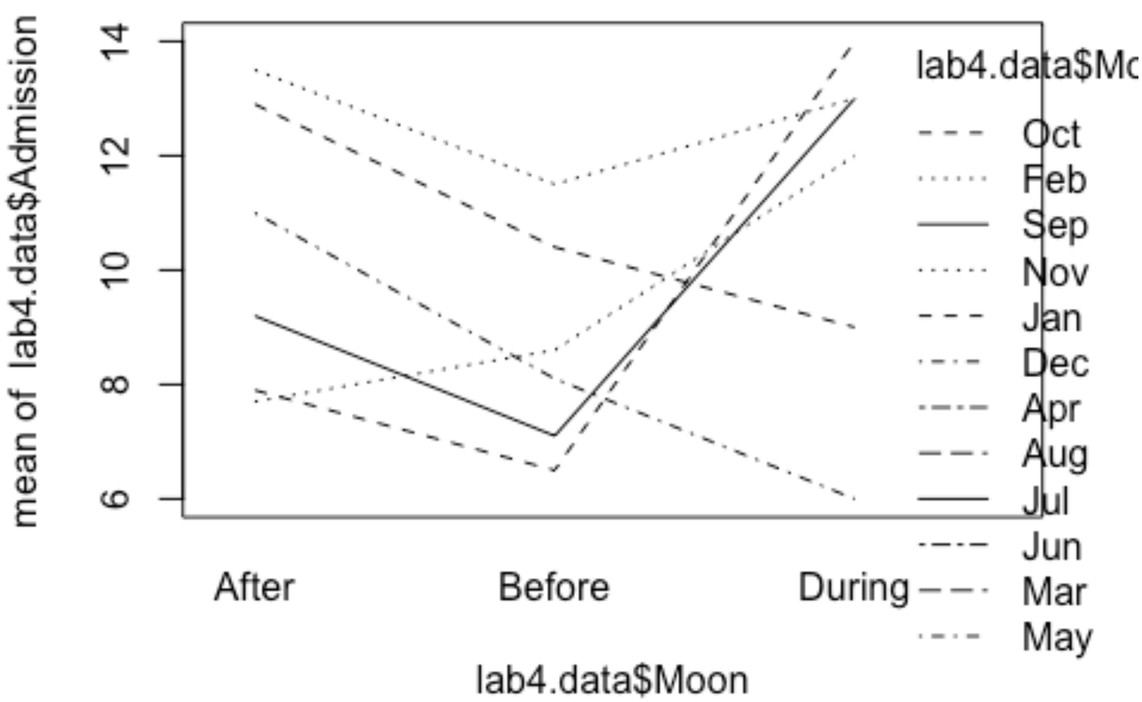
From the plots above, the line in Residuals vs Fitted plot isn't a perfect straight line but the residuals look normally distributed in the plot so the data has a constant variance. From Normal QQ plot, all the dots are on the straight linear line so the data is normal and we assume the data is randomlly collected, so the assumption is met.

```
anova(foo2)

## Analysis of Variance Table
##
## Response: Admission
##           Df Sum Sq Mean Sq F value Pr(>F)
## Moon       2 19.004  9.5022  1.4951 0.2703
## Month      5 32.971  6.5942  1.0376 0.4469
## Residuals 10 63.556  6.3556
```

From the two way anova test above, all the p values are greater than 0.05, so they are all not significant, but I still want to check if their interaction is sighificant so I will draw a interaction plot first.

```
interaction.plot(lab4.data$Moon, lab4.data$Month, lab4.data$Admission)
```



From the interaction plot, we can see they are not parallel to each other, so there must are some interactions in our model.

```
foo3<-aov(Admission~Moon*Month, data=lab4.data)
anova(foo3)

## Warning in anova.lm(foo3): ANOVA F-tests on an essentially perfect fit are
## unreliable

## Analysis of Variance Table
##
## Response: Admission
##            Df Sum Sq Mean Sq F value Pr(>F)
## Moon        2 19.004  9.5022
## Month       5 32.971  6.5942
## Moon:Month 10 63.556  6.3556
## Residuals   0  0.000
```

We don't have any residuals so the degree of freedom is 0, which means we can't comput p value, so we don't need this interaction term, which means there isn't any interaction for this data.

For the second Two-way anova test, I will create a new variable called Season:

```
##create new variable
lab4.data$Season=ifelse(lab4.data$Month %in% Winter, c("Winter"), c("Fall"))

head(lab4.data)

##   Month   Moon Admission Season
## 2   Sep Before       7.1   Fall
## 3   Oct Before       6.5   Fall
## 4   Nov Before       8.6   Fall
## 5   Dec Before       8.1 Winter
## 6   Jan Before      10.4 Winter
## 7   Feb Before      11.5 Winter
```

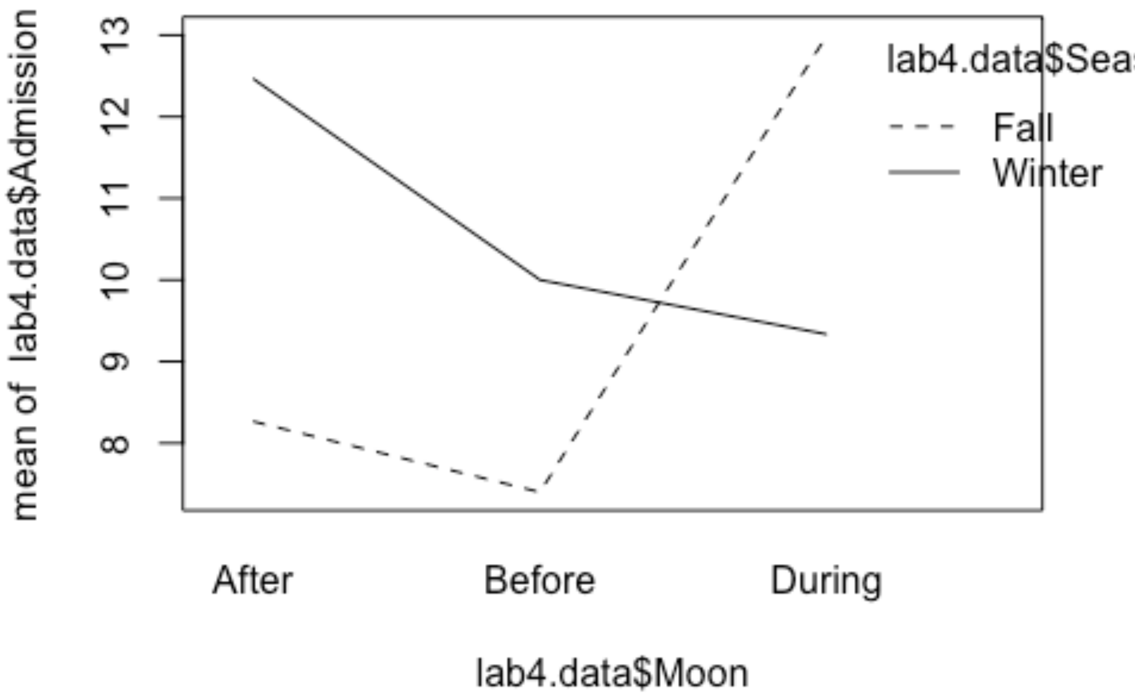
In the code below, I will do a two-way anova test with Moon and the new variable Season:

```
anova(foo4<-aov(Admission~as.factor(Season)+Moon, data=lab4.data))

## Analysis of Variance Table
##
## Response: Admission
##                   Df Sum Sq Mean Sq F value Pr(>F)
## as.factor(Season)  1  4.909  4.9089  0.7501 0.4010
## Moon               2 19.004  9.5022  1.4520 0.2673
## Residuals         14 91.618  6.5441
```

From the anova test above, these two variables p values are not significant, so I will draw a interaction plot to see which if theres an interaction between these to variables.

```
interaction.plot(lab4.data$Moon, lab4.data$Season, lab4.data$Admission)
```



From the interaction plot, we can see they are not parallel to each other, so there must are some interactions in our model and I can also see the After to Before level, Before and During level should also significant, so I need to do a TukeyHSD after checking interaction.

```
anova(foo5<-aov(Admission~as.factor(Season)*Moon, data=lab4.data))

## Analysis of Variance Table
##
## Response: Admission
##                        Df Sum Sq Mean Sq F value  Pr(>F)
## as.factor(Season)       1  4.909  4.9089  1.4816 0.24693
## Moon                    2 19.004  9.5022  2.8679 0.09594 .
## as.factor(Season):Moon  2 51.858 25.9289  7.8256 0.00668 **
## Residuals              12 39.760  3.3133
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

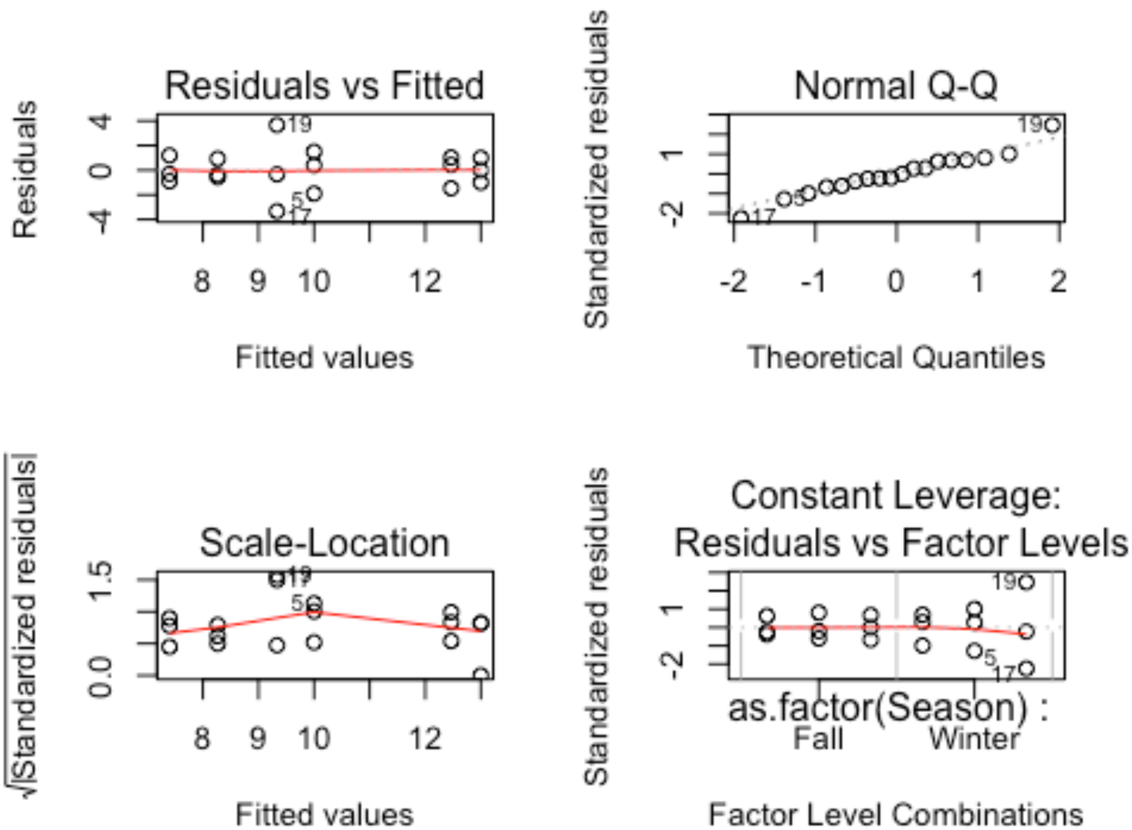
From the anova test above, the interaction term is significant so the interaction term is useful for this data.

Since the interaction term is significant to my data, I need to check if this model meet my assumption for anova:

Here is my hypotheses:

$H_0: \mu_1 = \mu_2 = \mu_3$ H_a: At least one μ differ from the rest

```
par(mfrow=c(2,2))
plot(foo5)
```




```
par(mfrow=c(1,1))
```

From the plots above, the line in Residuals vs Fitted plot has a linear straight line so the data has a constant variance. From Normal QQ plot, all the dots are on the straight linear line so the data is normal and we assume the data is randomlly collected, so the assumption is met.

Since it's a categorial data, I will use TukeyHSD test to see if my guessing was right.

```
TukeyHSD(foo5, which="as.factor(Season):Moon", ordered = TRUE)

##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##     factor levels have been ordered
##
## Fit: aov(formula = Admission ~ as.factor(Season) * Moon, data = lab4.data)
##
```

```
## $`as.factor(Season):Moon`
##                                  diff         lwr       upr     p adj
## Fall:After-Fall:Before      0.8666667 -4.12547278  5.858806 0.9903335
## Winter:During-Fall:Before   1.9333333 -3.05880611  6.925473 0.7794788
## Winter:Before-Fall:Before   2.6000000 -2.39213944  7.592139 0.5282500
## Winter:After-Fall:Before    5.0666667  0.07452722 10.058806 0.0459683
## Fall:During-Fall:Before     5.6000000  0.60786056 10.592139 0.0250968
## Winter:During-Fall:After    1.0666667 -3.92547278  6.058806 0.9759508
## Winter:Before-Fall:After    1.7333333 -3.25880611  6.725473 0.8441782
## Winter:After-Fall:After     4.2000000 -0.79213944  9.192139 0.1198782
## Fall:During-Fall:After      4.7333333 -0.25880611  9.725473 0.0668332
## Winter:Before-Winter:During 0.6666667 -4.32547278  5.658806 0.9971124
## Winter:After-Winter:During  3.1333333 -1.85880611  8.125473 0.3444931
## Fall:During-Winter:During   3.6666667 -1.32547278  8.658806 0.2084322
## Winter:After-Winter:Before  2.4666667 -2.52547278  7.458806 0.5791669
## Fall:During-Winter:Before   3.0000000 -1.99213944  7.992139 0.3864530
## Fall:During-Winter:After    0.5333333 -4.45880611  5.525473 0.9990001
```

From the TukeyHSD test above I can clearly see the Winter:After-Fall:Before and Fall:During-Fall:Before are significant, so these variables are affecting the hospital admissions.