

Hospital Admissions Project

ANOVA Models

Brady Fisher

Introduction

For this project, I will be using data from the MentalHealth dataset in the Stat2Data package, with the goal of testing 3 different anova models to see if there is a relationship between the population mean hospital admissions, moon phases(before, during or after the full moon) and the month/season of the year.

Loading the Data:

```
MentalHealth = read.csv("http://www.stat2.org/datasets/MentalHealth.csv")
```

I have been tasked with running the ANOVA tests on only the summer and winter months. (Excluding Fall and Spring Months)

Subset the Data:

```
myDataWinter = subset(MentalHealth, Month == "Dec" | Month == "Jan" | Month == "Feb")
myWinterData = data.frame(myDataWinter, Season = rep("Winter",9))
myDataSummer = subset(MentalHealth, Month == "Jun" | Month == "Jul" | Month == "Aug")
mySummerData = data.frame(myDataSummer, Season = rep("Summer",9))
myData = rbind(mySummerData, myWinterData)
```

With My Final Data being:

myData

##	Month	Moon	Admission	Season
## 1	Aug	Before	6.4	Summer
## 11	Jun	Before	11.7	Summer
## 12	Jul	Before	15.8	Summer
## 13	Aug	During	5.0	Summer
## 23	Jun	During	14.0	Summer
## 24	Jul	During	20.0	Summer
## 25	Aug	After	5.8	Summer
## 35	Jun	After	12.8	Summer
## 36	Jul	After	14.5	Summer
## 5	Dec	Before	8.1	Winter
## 6	Jan	Before	10.4	Winter
## 7	Feb	Before	11.5	Winter
## 17	Dec	During	6.0	Winter
## 18	Jan	During	9.0	Winter
## 19	Feb	During	13.0	Winter
## 29	Dec	After	11.0	Winter
## 30	Jan	After	12.9	Winter
## 31	Feb	After	13.5	Winter

Model 1: One-Way ANOVA

Choose the Model

I want to first see if there is a difference in hospital admissions for the different phases of the moon, so I will conduct a One-Way ANOVA test. With hypothesis:

H_0 : All three moon phases have the same population mean hospital admission rate

H_a : Not all three moon phases have the same population mean hospital admission rate

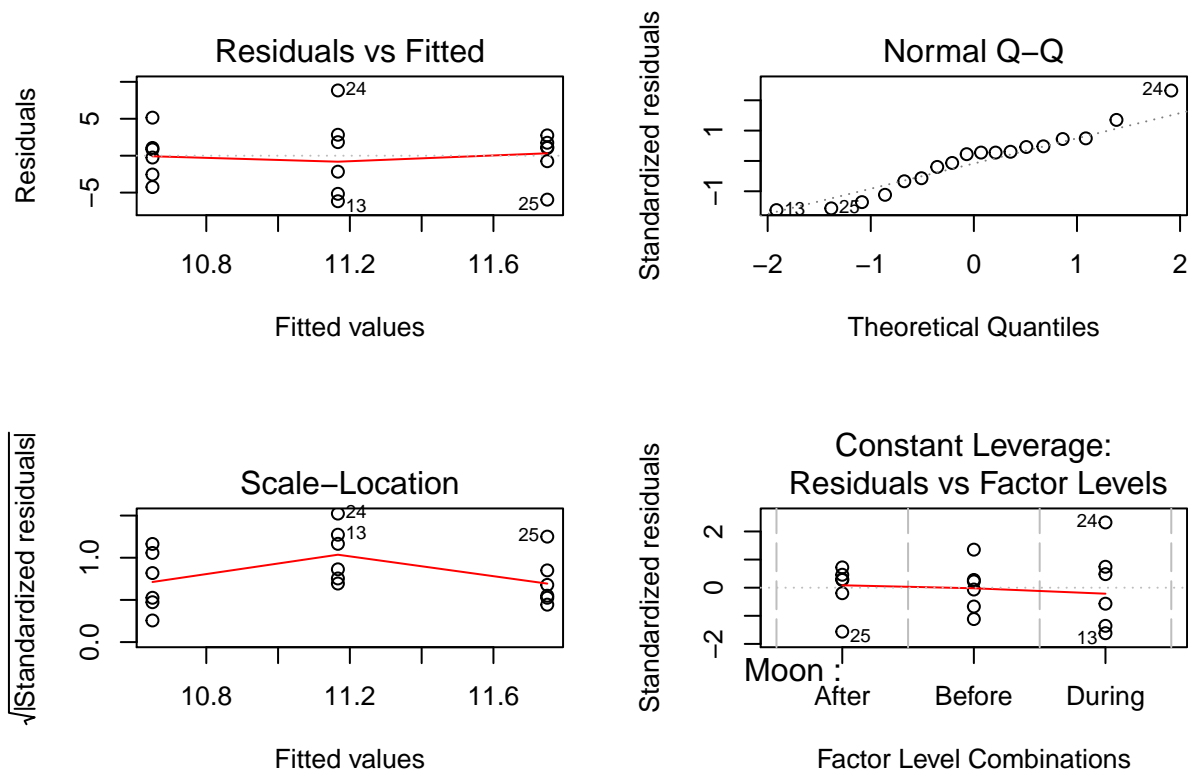
Fit the Model

```
m1 = aov(Admission ~ Moon, data = myData)
```

Assess the Model

```
par(mfrow = c(2, 2))
plot(m1)
library("car", lib.loc = "~/R/win-library/3.3")
```

```
## Warning: package 'car' was built under R version 3.3.2
```



```
outlierTest(m1)
```

```
##
## No Studentized residuals with Bonferonni p < 0.05
## Largest |rstudent|:
##      rstudent unadjusted p-value Bonferonni p
## 24 2.800104      0.014177      0.25519
```

```
which(cooks.distance(m1) > 0.5)
```

```
## named integer(0)
```

Now to use the One-Way ANOVA model, I must check the three assumptions of: 1.Independence 2.Constant variance 3.Normal Distribution

The first assumption of independence should be met from the study when the data was collected. It would also make sense that each sample would not dependent on another sample regardless of that samples moon phase. Thus the samples would be independent of each other.

The second assumption of constant variance is probably met. On the Residual vs Fitted plot the spread of the middle values, which corresponds to a moon phase of “During”, is a little wider, but this is solely caused by the point 24. I also tested for outliers and saw that there are no outliers or influential points in the data, so I should keep this point and assume constant variance.

The third assumption of Normal Distribution is probably met because the Normal Q-Q forms an approximately straight diagonal line, other than the last point (point 24), which I previously talked about. Also the ANOVA test gets more robust against the normality assumption as the number of observations increases and I have a decent amount of observations($n = 18$).

Use the Model

```
summary(m1)
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## Moon          2   3.63   1.817   0.105  0.901
## Residuals    15 260.82  17.388
```

From the ANOVA table I see that the test produced a F statistic of 0.105 which corresponds to a p-value of 0.901 which is way greater than 0.05. This leads me to the conclusion of not rejecting the null hypothesis meaning there is NOT strong enough evidence to say there is a difference in the population mean hospital admission rate in at least one of the moon phases. Since I did not find a difference I should not conduct a Tukey’s HSD test or any other comparisons test.

Model 2: First Two-Way ANOVA

Choose the Model

I want to correct for the potential fluctuations in the hospital admission that occur because of the time of year, so I will add the predictor Month into the model, and fit a two-way ANOVA model for Admission on Moon phase and Month of the year which is controlled for.

For a Two-Way ANOVA test there is an additive and an interaction model, but in this case I cannot use the interaction model because each factor level combination only has 1 observation. This means that I have the same number of parameters to estimate as the number of total observations, and thus there is no way to separate estimates of interaction and error. In other words there are no residual degrees of freedom, so the Mean Squares for the residual cannot be calculated(cannot divide by 0).

Fit the Model

```
m2 = aov(Admission ~ Moon + Month, data = myData)
```

Here month is working as a block in the model so I can account for any potential fluctuations in the hospital admissions cause by the time of the year.

Assess the Model

For this Model I was told to assume all the assumptions for a Two-Way ANOVA test were met.

Use the Model

```
summary(m2)
```

```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## Moon          2   3.63    1.82    0.465 0.641239
## Month         5 221.72   44.34   11.339 0.000728 ***
## Residuals    10  39.11    3.91
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The ANOVA table shows the F statistic for the Moon predictor is 0.465 and corresponds to a p-value of $0.641 > 0.05$. I also see that the F statistic for the Month predictor is 11.339 and corresponds to a p-value of $0.000728 < 0.05$. This means that the effect of the moon phase is not significant to the population mean hospital admission rate, but the effect of the Month of the year is significant after adjusting for the effect of the moon phase.

Model 3: Second Two-Way ANOVA

Choose the Model

Now I want to now use a Two-way ANOVA model that controls for the season of the year instead of controlling for the month of the year like in Model 2. This way I will be able to fit the interaction model, since here all the level combinations will have 3 observations.

Fit the Model

The interaction and the additive models are:

```
m3 = aov(Admission ~ Moon * Season, data = myData)
m3a = aov(Admission ~ Moon + Season, data = myData)
```

Assess the Model

Again, for this Model I was told to assume all the assumptions for a Two-Way ANOVA test were met.

Use the Model

```
summary(m3)
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## Moon          2   3.63   1.817   0.093  0.912
## Season        1   6.24   6.242   0.319  0.583
## Moon:Season    2  19.54   9.771   0.499  0.619
## Residuals    12 235.04  19.587
```

Here I see that the F statistic for the interaction term Moon:Season is 0.499 and corresponds to a p-value of $0.619 > 0.05$. This means that the interaction term between the phase of the moon and the season is not significant. So the effect of the moon phase on hospital admission could be the same across the summer and winter months. Thus my final model will be the additive model.

I can determine which pair(s) of moon phases are significantly different directly through the TukeyHSD function.

```
TukeyHSD(m3a, which = "Moon")
```

```
##      Tukey multiple comparisons of means
##      95% family-wise confidence level
##
## Fit: aov(formula = Admission ~ Moon + Season, data = myData)
##
## $Moon
##              diff          lwr          upr          p adj
## Before-After -1.1000000 -7.543750  5.343750  0.8965960
## During-After -0.5833333 -7.027084  5.860417  0.9695938
## During-Before  0.5166667 -5.927084  6.960417  0.9760556
```

From this result I see that the comparison between Before and After moon phase has an interval of $(-7.544, 5.344)$ and a p-value of 0.8966. The comparison between During and After moon phase has an interval of $(-7.027, 5.86)$ and a p-value of 0.9696. The comparison between During and Before moon phase has an interval of $(-5.927, 6.960)$ and a p-value of 0.9761.

Since all 3 comparisons have large p-values much greater than 0.05 and have intervals that include 0 the moon phase does not help predict hospital admissions differently for the different seasons (winter and summer). So I cannot say which phase of the moon or season has the highest hospital admissions.