Advanced Stats

Sergio

October 28, 2016

I am loading all the libraries that will be utilized

in this analysis.

```
library(car)
library(Rcmdr)
```

```
## Loading required package: splines

## Loading required package: RcmdrMisc

## Loading required package: sandwich

## The Commander GUI is launched only in interactive sessions
```

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following object is masked from 'package:car':
##
##     recode

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ggplot2)
library(e1071)
library(reshape)
```

```
##
## Attaching package: 'reshape'

## The following object is masked from 'package:dplyr':
##
##     rename
```

```
library(reshape2)
```

```
##
## Attaching package: 'reshape2'

## The following objects are masked from 'package:reshape':
##
##     colsplit, melt, recast
```

The files that will be analyzed are being loaded and read.

```
read.csv("ERWaiting.csv")
```

```
##    day   Main Satellite.1 Satellite.2 Satellite.3
## 1    1 120.08       30.75       75.86       54.05
## 2    2  81.90       61.83       37.88       38.82
## 3    3  78.79       26.40       68.73       36.85
## 4    4  63.83       53.84       51.08       32.83
## 5    5  79.77       72.30       50.21       52.94
## 6    6  47.94       53.09       58.47       34.13
## 7    7  79.88       27.67       86.29       69.37
## 8    8  48.63       52.46       62.90       78.52
## 9    9  55.43       10.64       44.84       55.95
## 10  10  64.06       53.50       64.17       49.61
## 11  11  64.99       37.28       50.68       66.40
## 12  12  53.82       34.31       47.97       76.06
## 13  13  62.43       66.00       60.57       11.37
## 14  14  65.07        8.99       58.37       83.51
## 15  15  81.02       29.75       30.40       39.17
```

```
waiting <- read.csv("ERWaiting.csv")
```

The following runs a summary statistical printout. We are able to analyze median, mean, minimum, maximum, first and Third quartiles. This a very good way to get a feel of what the data looks like.

```
summary(waiting)
```

```
##       day             Main          Satellite.1     Satellite.2
##  Min.   : 1.0   Min.   : 47.94   Min.   : 8.99   Min.   :30.40
##  1st Qu.: 4.5   1st Qu.: 58.93   1st Qu.:28.71   1st Qu.:49.09
##  Median : 8.0   Median : 64.99   Median :37.28   Median :58.37
##  Mean   : 8.0   Mean   : 69.84   Mean   :41.25   Mean   :56.56
##  3rd Qu.:11.5   3rd Qu.: 79.83   3rd Qu.:53.67   3rd Qu.:63.53
##  Max.   :15.0   Max.   :120.08   Max.   :72.30   Max.   :86.29
##   Satellite.3
##  Min.   :11.37
##  1st Qu.:37.84
##  Median :52.94
##  Mean   :51.97
##  3rd Qu.:67.89
##  Max.   :83.51
```

In order to analyze the data and run different methods such as ANOVA we must group the data in a way that allows us to run such analisis. These are the first lines of code in process to group the data in vectors.

```
day <- c(waiting$day)
main <- c(waiting$Main)
s1 <- c(waiting$Satellite.1)
s2 <- c(waiting$Satellite.2)
s3 <- c(waiting$Satellite.3)
```

I get the vectors and create a data frame in order to prepare the data for analysis. I also want to run the data frame to view the data and the output.

```
gr <- data.frame(cbind(day,main,s1,s2,s3))
gr
```

```
##    day   main    s1    s2    s3
## 1    1 120.08 30.75 75.86 54.05
## 2    2  81.90 61.83 37.88 38.82
## 3    3  78.79 26.40 68.73 36.85
## 4    4  63.83 53.84 51.08 32.83
## 5    5  79.77 72.30 50.21 52.94
## 6    6  47.94 53.09 58.47 34.13
## 7    7  79.88 27.67 86.29 69.37
## 8    8  48.63 52.46 62.90 78.52
## 9    9  55.43 10.64 44.84 55.95
## 10  10  64.06 53.50 64.17 49.61
## 11  11  64.99 37.28 50.68 66.40
## 12  12  53.82 34.31 47.97 76.06
## 13  13  62.43 66.00 60.57 11.37
## 14  14  65.07  8.99 58.37 83.51
## 15  15  81.02 29.75 30.40 39.17
```

```
summary(gr)
```

```
##       day             main             s1              s2
##  Min.   : 1.0   Min.   : 47.94   Min.   : 8.99   Min.   :30.40
##  1st Qu.: 4.5   1st Qu.: 58.93   1st Qu.:28.71   1st Qu.:49.09
##  Median : 8.0   Median : 64.99   Median :37.28   Median :58.37
##  Mean   : 8.0   Mean   : 69.84   Mean   :41.25   Mean   :56.56
##  3rd Qu.:11.5   3rd Qu.: 79.83   3rd Qu.:53.67   3rd Qu.:63.53
##  Max.   :15.0   Max.   :120.08   Max.   :72.30   Max.   :86.29
##        s3
##  Min.   :11.37
##  1st Qu.:37.84
##  Median :52.94
##  Mean   :51.97
##  3rd Qu.:67.89
##  Max.   :83.51
```

This code prepares the data in a way that allows for an ANOVA test can be conducted. This line of code stacks the data in two rows. A summary is also printed out to view the data in prepared form.

```
stak <- stack(gr)
stak
```

```
##    values  ind
## 1    1.00  day
## 2    2.00  day
## 3    3.00  day
## 4    4.00  day
## 5    5.00  day
## 6    6.00  day
## 7    7.00  day
## 8    8.00  day
## 9    9.00  day
## 10  10.00  day
## 11  11.00  day
## 12  12.00  day
## 13  13.00  day
## 14  14.00  day
## 15  15.00  day
## 16 120.08 main
## 17  81.90 main
## 18  78.79 main
## 19  63.83 main
## 20  79.77 main
## 21  47.94 main
## 22  79.88 main
## 23  48.63 main
## 24  55.43 main
## 25  64.06 main
## 26  64.99 main
## 27  53.82 main
## 28  62.43 main
## 29  65.07 main
## 30  81.02 main
## 31  30.75   s1
## 32  61.83   s1
## 33  26.40   s1
## 34  53.84   s1
## 35  72.30   s1
## 36  53.09   s1
## 37  27.67   s1
## 38  52.46   s1
## 39  10.64   s1
## 40  53.50   s1
## 41  37.28   s1
## 42  34.31   s1
## 43  66.00   s1
## 44   8.99   s1
## 45  29.75   s1
## 46  75.86   s2
```

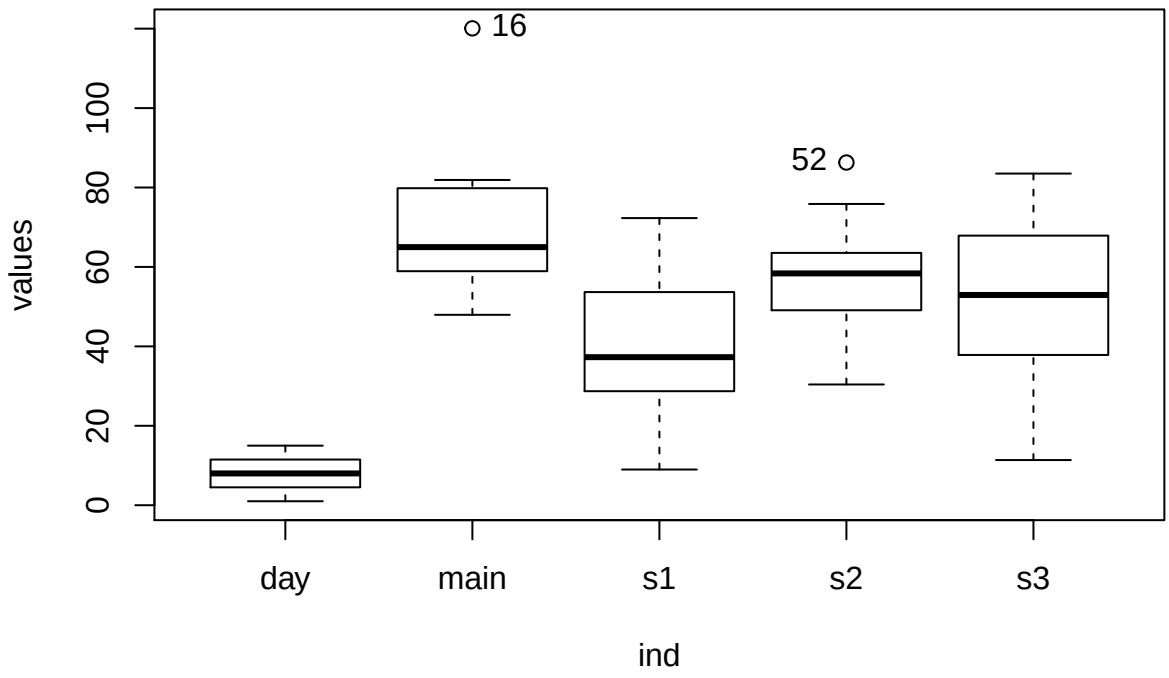
```
## 47  37.88   s2
## 48  68.73   s2
## 49  51.08   s2
## 50  50.21   s2
## 51  58.47   s2
## 52  86.29   s2
## 53  62.90   s2
## 54  44.84   s2
## 55  64.17   s2
## 56  50.68   s2
## 57  47.97   s2
## 58  60.57   s2
## 59  58.37   s2
## 60  30.40   s2
## 61  54.05   s3
## 62  38.82   s3
## 63  36.85   s3
## 64  32.83   s3
## 65  52.94   s3
## 66  34.13   s3
## 67  69.37   s3
## 68  78.52   s3
## 69  55.95   s3
## 70  49.61   s3
## 71  66.40   s3
## 72  76.06   s3
## 73  11.37   s3
## 74  83.51   s3
## 75  39.17   s3
```

```
summary(stak)
```

```
##      values         ind
##  Min.   :  1.00   day :15
##  1st Qu.: 27.04   main:15
##  Median : 50.68   s1  :15
##  Mean   : 45.53   s2  :15
##  3rd Qu.: 64.11   s3  :15
##  Max.   :120.08
```

A boxplot graph is created here to show the way that shows how the means of each vector compare. This is a good way to view any outliers and to see if the data is normal or homogeneous.

```
Boxplot(values~ind, data=stak, id.method="y")
```




```
## [1] "16" "52"
```

Here we run an ANOVA analisis and see the results.It clearly shows that by having such a low F-value the data values are not the same. we can conclude that the means are not similar in each satelite location. A Turkey analysis is also conducted to get a better view of the data and hypothesis. The hypothesis was created with the Ho: to indicate the data to be equal and the alternative to be different. Here we clearly see that the Ho is rejected and the alternative accepted.

```
Anova_results <- aov(values~ind,data=stak)
summary(Anova_results)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## ind          4  32716    8179    30.5 1.07e-14 ***
## Residuals   70  18773     268
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Anova_TukeyR <-TukeyHSD(Anova_results)
Anova_TukeyR
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
```

```
## Fit: aov(formula = values ~ ind, data = stak)
##
## $ind
##                diff        lwr        upr     p adj
## main-day  61.842667  45.098250  78.587084 0.0000000
## s1-day    33.254000  16.509583  49.998417 0.0000045
## s2-day    48.561333  31.816916  65.305750 0.0000000
## s3-day    43.972000  27.227583  60.716417 0.0000000
## s1-main  -28.588667 -45.333084 -11.844250 0.0000895
## s2-main  -13.281333 -30.025750   3.463084 0.1841192
## s3-main  -17.870667 -34.615084  -1.126250 0.0306450
## s2-s1     15.307333  -1.437084  32.051750 0.0892830
## s3-s1     10.718000  -6.026417  27.462417 0.3861853
## s3-s2     -4.589333 -21.333750  12.155084 0.9390637
```

The levens test is conducted here in order to test the residuals of the ANOVA and see if the data is either Normal or Homogeneous. In this case the test shows that that the data is not Homogenous.

```
leveneTest(values ~ ind, data = stak)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value   Pr(>F)
## group  4  3.8562 0.006879 **
##       70
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```