

# Homework 2-Descriptive Statistics and ANOVA of ER Waiting time

*An implementation in R Markdown*

*Viviana Berry*

*2016-10-28*

## *Descriptive Statistics:*

Our variables, Main, Satellite.1, Satellite.2 and Satellite.3, seem to have very different means but when we take into consideration their variability all of them seem to overlap at least with one other variable.

Therefore, we perform a descriptive statistics analysis considering other factors as shown in the table below:

```
print(summary.data)
```

##	day	Main	Satellite.1	Satellite.2	Satellite.3
## mean	8.000000	69.842667	41.25400000	56.561333	51.9720000
## sd	4.472136	18.202192	19.39010160	14.330152	20.2060925
## var	20.000000	331.319792	375.97604000	205.353255	408.2861743
## Min	1.000000	47.940000	8.99000000	30.400000	11.3700000
## q1	4.500000	58.930000	28.71000000	49.090000	37.8350000
## median	8.000000	64.990000	37.28000000	58.370000	52.9400000
## q3	11.500000	79.825000	53.67000000	63.535000	67.8850000
## Max	15.000000	120.080000	72.30000000	86.290000	83.5100000

```
## Kurt    -1.441333    1.239497   -1.28180663   -0.525495   -0.9922604
## Skew     0.000000    1.154559   -0.09562142    0.198051   -0.1217972
```

## *Analysis of Variance:*

In order to test this assumption, this hypothesis testing needs to be done:

Ho: All averages of ER waiting are the same  
 H1: At least one group has a different average

```
print(fit1.aov.summary)
```

```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## as.factor(variable)  3    6312   2104.1     6.372 0.000859 ***
## Residuals          56   18493     330.2
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Results;

With 95% Confidence level, we can reject Ho since P\_value is .00086 which implies that not all ER waiting time are the same.

## *Analysis of Tukey Test:*

With a Confidence Interval of .05, the only two pairs that differ from each other are Satellite.1–Main and Satellite.3–Main. All the other pairings P\_values are larger than the confidence interval; therefore, we can conclude that those variables are similar.

```
print(erdata.tukey)

## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = value ~ as.factor(variable), data = erdata.melt)
##
## $`as.factor(variable)`
##
```

	diff	lwr	upr	p adj
## Satellite.1-Main	-28.588667	-46.158986	-11.0183478	0.0003814
## Satellite.2-Main	-13.281333	-30.851652	4.2889855	0.1997139
## Satellite.3-Main	-17.870667	-35.440986	-0.3003478	0.0447878
## Satellite.2-Satellite.1	15.307333	-2.262986	32.8776522	0.1086697
## Satellite.3-Satellite.1	10.718000	-6.852319	28.2883188	0.3785196
## Satellite.3-Satellite.2	-4.589333	-22.159652	12.9809855	0.8998737

## *Analysis of Levene Test:*

Levene's test p-value is above .05, and therefore we can reject the null hypothesis that the variance in all groups are the same.

```
print(ER.levene)

## Levene's Test for Homogeneity of Variance (center = median)
##      Df F value Pr(>F)
## group 3  0.8201 0.4883
##      56
```

## *Normality of Data:*

The assumptions on our analysis are based on that the data is normal distributed. From the descriptive statistics, all of the hospitals have a skewness between -1 and 1 except the Main hospital where the value is slightly above 1. Finally, the plot

shows that the sample follows roughly a straight line, so there is enough evidence to conclude the data is normal.

```
print(qqp)
```

