

# Comparison of Datasets by ANOVA

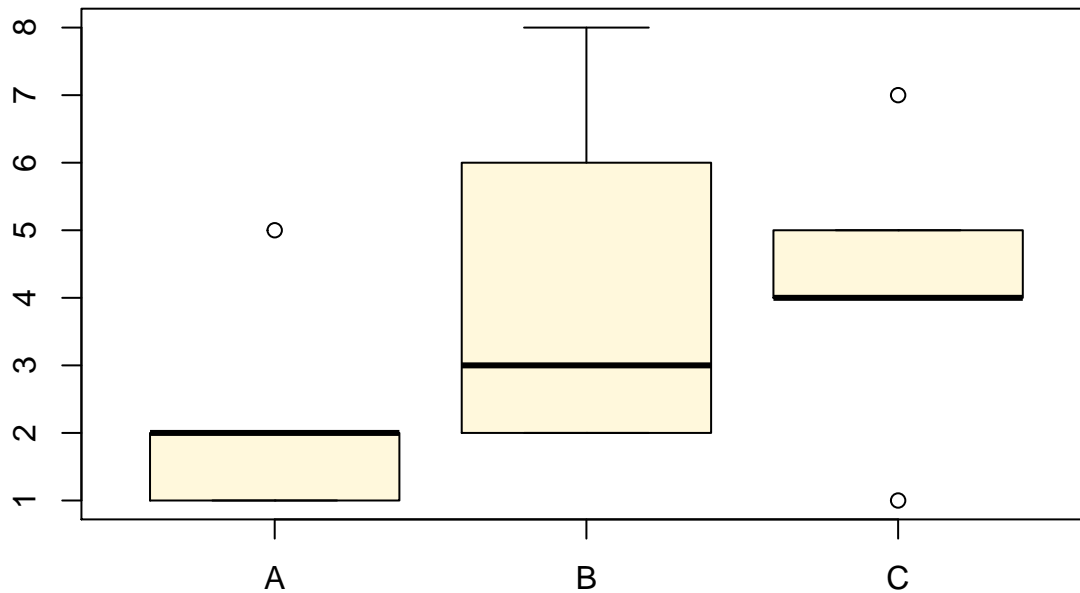
02 April, 2016

```
x <- read.csv("~/Documents/GitHub/ANOVA/data/data.csv", as.is=TRUE, header=TRUE)
units <- ""
```

The following data is to be assessed as to whether there are significant differences between the individual groups.

```
##   A B C
## 1 1 2 1
## 2 2 2 7
## 3 5 3 4
## 4 2 6 4
## 5 5 1 8 5
```

This can be displayed in a Boxplot visualisation. Circles, if present, represent outliers.



An ANOVA assessment of the data is:

```
##           Df Sum Sq Mean Sq F value Pr(>F)
## ind         2  13.33   6.667    1.37  0.291
## Residuals   12  58.40   4.867
```

If the “Pr(>F)” value is less than 0.05, it indicates that there is a significant difference between the means of the data sets but doesn’t indicate which.

Looking at the individual differences, using Tukey’s HSD criteria, gives an indication of which means differ significantly from each other. A “p adj” value of less than 0.05 indicates a significant difference.

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

```
##
## Fit: aov(formula = values ~ ind, data = xs)
##
## $ind
##          diff          lwr          upr          p adj
## B-A 2.000000e+00 -1.722283  5.722283  0.3556618
## C-A 2.000000e+00 -1.722283  5.722283  0.3556618
## C-B 8.881784e-16 -3.722283  3.722283  1.0000000
```