Hypothesis Testing ANOVA

Abhirup Sen

14/05/2021

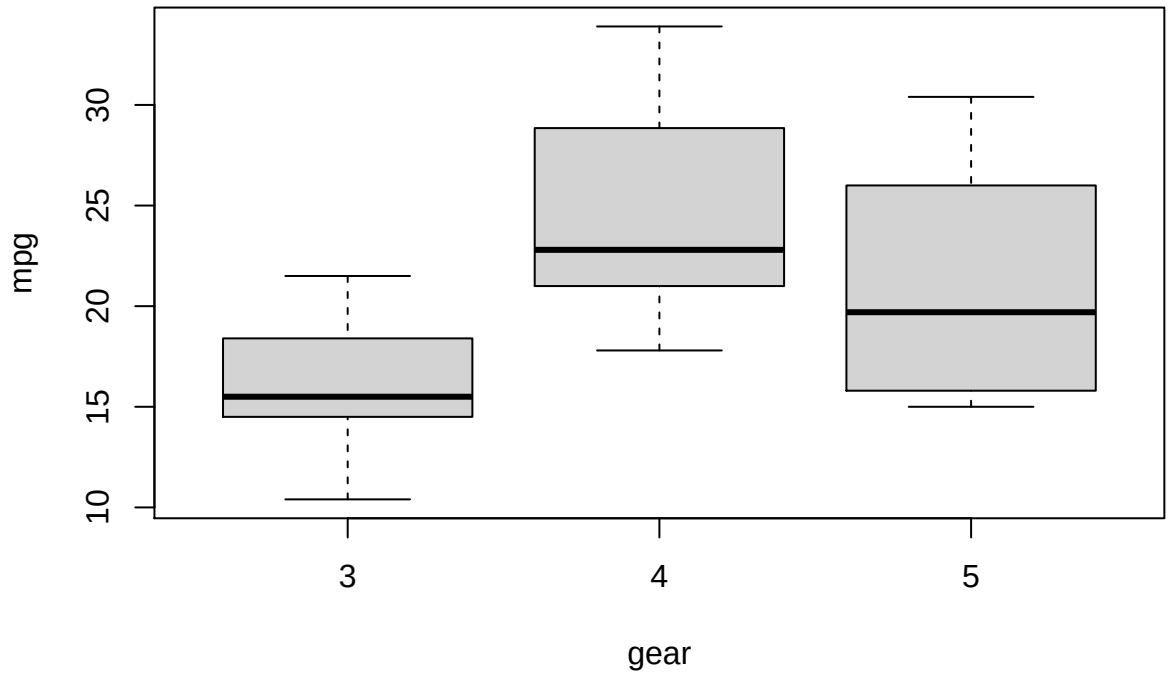
#One way anova

##step 1 : Set up the nul hypothesis and the alternative hypothesis ##H0 = mu0=mu1=mu2 (there is no difference in avg mpg for different gear) ##H1 = Not all means are equal

```
mtcars
```

```
##                     mpg cyl  disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4          21.0   6 160.0 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag      21.0   6 160.0 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710         22.8   4 108.0  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive     21.4   6 258.0 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout  18.7   8 360.0 175 3.15 3.440 17.02  0  0    3    2
## Valiant            18.1   6 225.0 105 2.76 3.460 20.22  1  0    3    1
## Duster 360         14.3   8 360.0 245 3.21 3.570 15.84  0  0    3    4
## Merc 240D          24.4   4 146.7  62 3.69 3.190 20.00  1  0    4    2
## Merc 230           22.8   4 140.8  95 3.92 3.150 22.90  1  0    4    2
## Merc 280           19.2   6 167.6 123 3.92 3.440 18.30  1  0    4    4
## Merc 280C          17.8   6 167.6 123 3.92 3.440 18.90  1  0    4    4
## Merc 450SE         16.4   8 275.8 180 3.07 4.070 17.40  0  0    3    3
## Merc 450SL         17.3   8 275.8 180 3.07 3.730 17.60  0  0    3    3
## Merc 450SLC        15.2   8 275.8 180 3.07 3.780 18.00  0  0    3    3
## Cadillac Fleetwood 10.4   8 472.0 205 2.93 5.250 17.98  0  0    3    4
## Lincoln Continental 10.4  8 460.0 215 3.00 5.424 17.82  0  0    3    4
## Chrysler Imperial  14.7   8 440.0 230 3.23 5.345 17.42  0  0    3    4
## Fiat 128           32.4   4  78.7  66 4.08 2.200 19.47  1  1    4    1
## Honda Civic        30.4   4  75.7  52 4.93 1.615 18.52  1  1    4    2
## Toyota Corolla     33.9   4  71.1  65 4.22 1.835 19.90  1  1    4    1
## Toyota Corona      21.5   4 120.1  97 3.70 2.465 20.01  1  0    3    1
## Dodge Challenger   15.5   8 318.0 150 2.76 3.520 16.87  0  0    3    2
## AMC Javelin        15.2   8 304.0 150 3.15 3.435 17.30  0  0    3    2
## Camaro Z28         13.3   8 350.0 245 3.73 3.840 15.41  0  0    3    4
## Pontiac Firebird   19.2   8 400.0 175 3.08 3.845 17.05  0  0    3    2
## Fiat X1-9          27.3   4  79.0  66 4.08 1.935 18.90  1  1    4    1
## Porsche 914-2      26.0   4 120.3  91 4.43 2.140 16.70  0  1    5    2
## Lotus Europa       30.4   4  95.1 113 3.77 1.513 16.90  1  1    5    2
## Ford Pantera L     15.8   8 351.0 264 4.22 3.170 14.50  0  1    5    4
## Ferrari Dino       19.7   6 145.0 175 3.62 2.770 15.50  0  1    5    6
## Maserati Bora      15.0   8 301.0 335 3.54 3.570 14.60  0  1    5    8
## Volvo 142E         21.4   4 121.0 109 4.11 2.780 18.60  1  1    4    2
```

```r
#See the variance between the groups and within the groups
boxplot(mtcars$mpg~factor(mtcars$gear),xlab="gear",ylab="mpg")
```



##Step 2: Calculate the Test Statistics using anova function

```r
mtcars.aov <- aov(mtcars$mpg ~ factor(mtcars$gear)) # factor makes it a charactre variable and there by
summary(mtcars.aov)
```

```
##                      Df Sum Sq Mean Sq F value   Pr(>F)
## factor(mtcars$gear)  2  483.2  241.62    10.9 0.000295 ***
## Residuals           29  642.8   22.17
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

##Step 3: Calculate F Critical Value ##for 0.05 significance level , critical value = alpha = 0.05

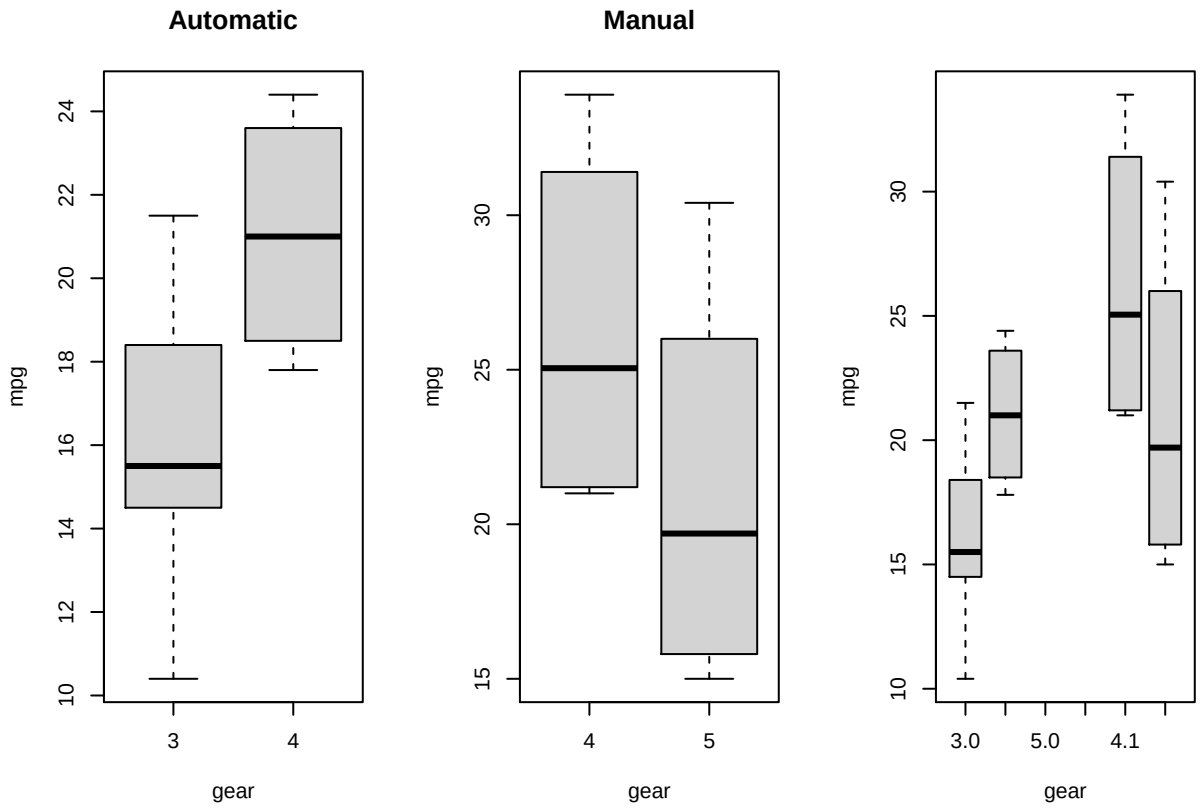
##Step 4 : Compare test statistics with F critical value and conclude test ##p < alpha - > Reject Hypothesis

##Conclusion : Gear and mpg has strong relation.

Two way anova .

##See the variance between the groups and within the groups

```
par(mfrow=c(1,3))
boxplot(mtcars$mpg~mtcars$gear, subset =(mtcars$am == 0),xlab = "gear",ylab="mpg",main="Automatic" )
boxplot(mtcars$mpg~mtcars$gear, subset =(mtcars$am == 1),xlab = "gear",ylab="mpg",main="Manual" )
boxplot(mtcars$mpg~factor(mtcars$gear)*factor(mtcars$am),xlab = "gear",ylab="mpg")
```



##step 1 : set up NULL hypothesis and Alternative Hypothesis ##H0 = mu0=mu1=mu2 (there is no difference in avg mpg for different gear) ##H1 = Not all means are equal

##Step 2 : Test Two way anova using anova function

```
mtcars.anova2 <- aov(mtcars$mpg~factor(mtcars$gear)*factor(mtcars$am))
summary(mtcars.anova2)
```

```
##                      Df Sum Sq Mean Sq F value   Pr(>F)
## factor(mtcars$gear)   2  483.2  241.62  11.869 0.000185 ***
## factor(mtcars$am)     1   72.8   72.80   3.576 0.069001 .
## Residuals            28  570.0   20.36
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

##Step 3 : Calculate F Crticical value ## for 0.05 significance level, critical Value = alpha = 0.05

Step 4 : Compare test statistics with F Critical Value and conclude test

p< alpha for gear but not for transmission type

So mpg has strong relation with gear but not with transmission type #Conclusion : gear and mpg mean has different value which mean they have string relationship '

#Two Way ANOVA - POST HOC TEST

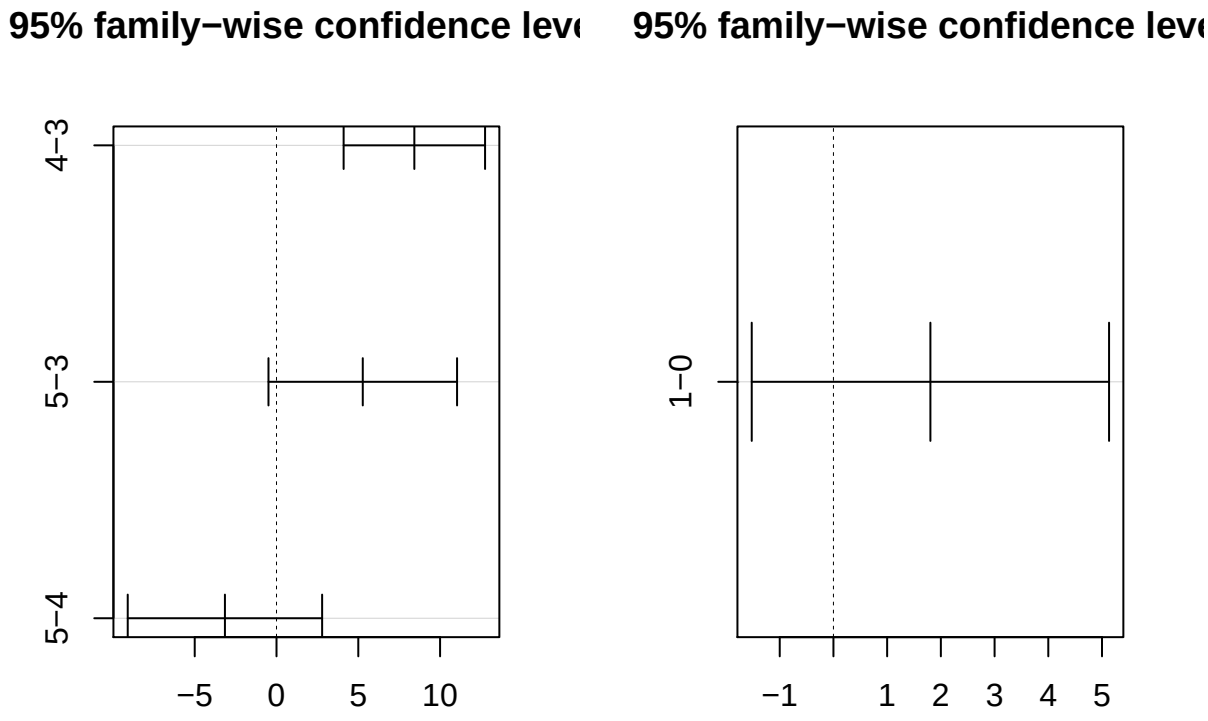
once the null hypothesis is rejected, the question is what treatment differ. Do post hoc analysis using Tukey Honest Significant Different plot

```
TukeyHSD(mtcars.anova2)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = mtcars$mpg ~ factor(mtcars$gear) * factor(mtcars$am))
##
## $`factor(mtcars$gear)`
##          diff        lwr       upr     p adj
## 4-3  8.426667  4.1028616 12.750472 0.0001301
## 5-3  5.273333 -0.4917401 11.038407 0.0779791
## 5-4 -3.153333 -9.0958350  2.789168 0.3999532
##
## $`factor(mtcars$am)`
##         diff       lwr     upr     p adj
## 1-0 1.805128 -1.521483 5.13174 0.2757926
```

Result indicates that the relation between gear and mpg is significant but relation between mpg and transmission type is not significant

```
par(mfrow=c(1,2))
plot(TukeyHSD(mtcars.anova2))
```



Result indicates that gear 4 and gear 3 has highest difference of mean for mpg.