

# Regression Models Motor Trend Analysis

*Qiang Wu*

*December 12, 2015*

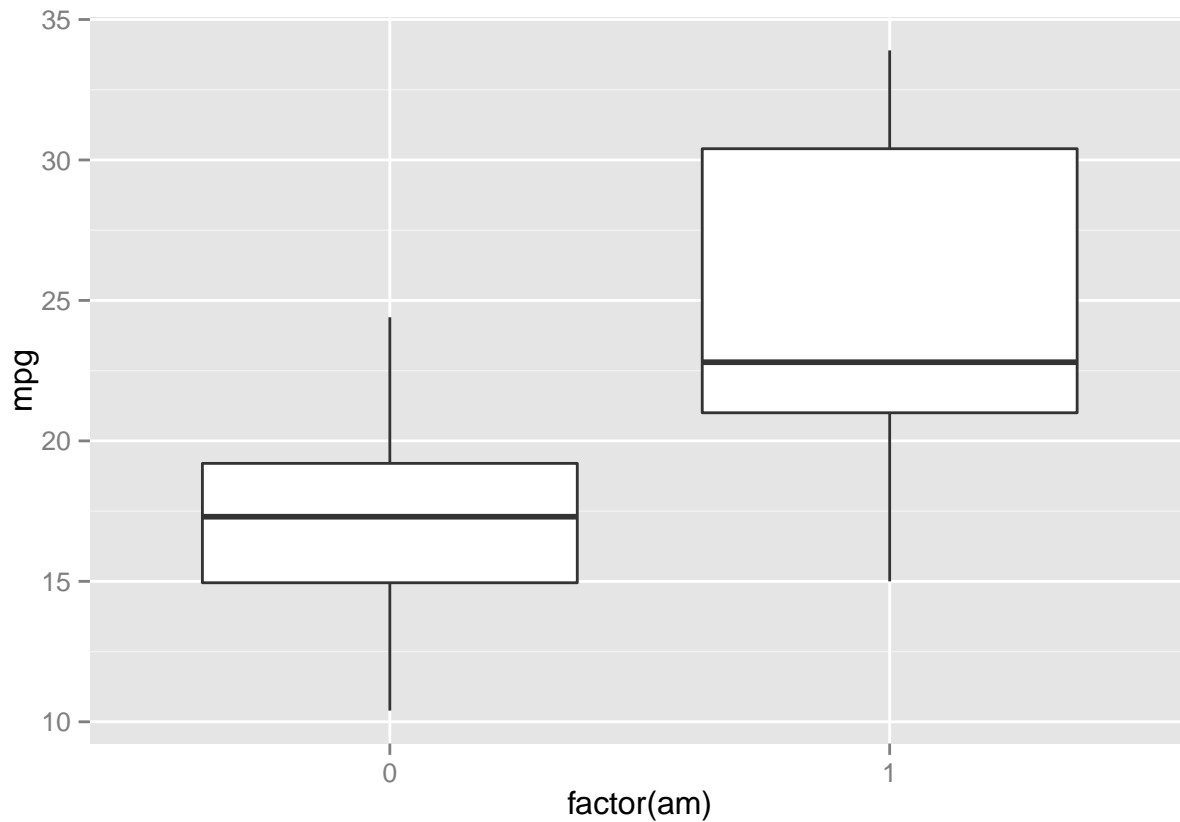
1. load data mtcars

```
library(ggplot2)
data("mtcars")
summary(mtcars)
```

```
##      mpg          cyl        disp         hp
##  Min.   :10.40   Min.    :4.000   Min.    : 71.1   Min.    : 52.0
## 1st Qu.:15.43   1st Qu.:4.000   1st Qu.:120.8   1st Qu.: 96.5
## Median :19.20   Median :6.000   Median :196.3   Median :123.0
## Mean   :20.09   Mean    :6.188   Mean    :230.7   Mean    :146.7
## 3rd Qu.:22.80   3rd Qu.:8.000   3rd Qu.:326.0   3rd Qu.:180.0
## Max.   :33.90   Max.    :8.000   Max.    :472.0   Max.    :335.0
##      drat          wt        qsec         vs
##  Min.   :2.760   Min.    :1.513   Min.    :14.50   Min.    :0.0000
## 1st Qu.:3.080   1st Qu.:2.581   1st Qu.:16.89   1st Qu.:0.0000
## Median :3.695   Median :3.325   Median :17.71   Median :0.0000
## Mean   :3.597   Mean    :3.217   Mean    :17.85   Mean    :0.4375
## 3rd Qu.:3.920   3rd Qu.:3.610   3rd Qu.:18.90   3rd Qu.:1.0000
## Max.   :4.930   Max.    :5.424   Max.    :22.90   Max.    :1.0000
##      am          gear        carb
##  Min.   :0.0000   Min.    :3.000   Min.    :1.000
## 1st Qu.:0.0000   1st Qu.:3.000   1st Qu.:2.000
## Median :0.0000   Median :4.000   Median :2.000
## Mean   :0.4062   Mean    :3.688   Mean    :2.812
## 3rd Qu.:1.0000   3rd Qu.:4.000   3rd Qu.:4.000
## Max.   :1.0000   Max.    :5.000   Max.    :8.000
```

2. First let's plot the mpg vs the auto/manual

```
ggplot(mtcars, aes(factor(am), mpg)) + geom_boxplot()
```



3. From the plot, we can clearly see there is mpg difference between mpg in auto and manual cars

```
median(mtcars[mtcars$am==0,]$mpg)
```

```
## [1] 17.3
```

```
median(mtcars[mtcars$am==1,]$mpg)
```

```
## [1] 22.8
```

4. Try to run regression.

```
fit<-lm(mpg~factor(am), mtcars)
summary(fit)
```

```
##
## Call:
## lm(formula = mpg ~ factor(am), data = mtcars)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -9.3923 -3.0923 -0.2974  3.2439  9.5077
##
```

```
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  17.147      1.125  15.247 1.13e-15 ***
## factor(am)1   7.245      1.764   4.106 0.000285 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.902 on 30 degrees of freedom
## Multiple R-squared:  0.3598, Adjusted R-squared:  0.3385
## F-statistic: 16.86 on 1 and 30 DF,  p-value: 0.000285
```

5. From fit, we can see there is 7.2449393 difference between auto/manual cars. It's also greater than  $2 * 1.7644216$  And p value is very small, so this means this slop is significant.

6. play with mtcars more, if we consider all information

```
summary(lm(mpg~., mtcars))
```

```
##
## Call:
## lm(formula = mpg ~ ., data = mtcars)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.4506 -1.6044 -0.1196  1.2193  4.6271
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) 12.30337   18.71788   0.657  0.5181
## cyl         -0.11144    1.04502  -0.107  0.9161
## disp          0.01334    0.01786   0.747  0.4635
## hp          -0.02148    0.02177  -0.987  0.3350
## drat          0.78711    1.63537   0.481  0.6353
## wt          -3.71530    1.89441  -1.961  0.0633 .
## qsec          0.82104    0.73084   1.123  0.2739
## vs           0.31776    2.10451   0.151  0.8814
## am           2.52023    2.05665   1.225  0.2340
## gear          0.65541    1.49326   0.439  0.6652
## carb        -0.19942    0.82875  -0.241  0.8122
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.65 on 21 degrees of freedom
## Multiple R-squared:  0.869, Adjusted R-squared:  0.8066
## F-statistic: 13.93 on 10 and 21 DF,  p-value: 3.793e-07
```

We can see no parameter is significant any more. But seems weight (wt) is more important than others.

7. If we think about am and wt only.

```
summary(lm(mpg~factor(am)+wt, mtcars))
```

```
##
## Call:
## lm(formula = mpg ~ factor(am) + wt, data = mtcars)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4.5295 -2.3619 -0.1317  1.4025  6.8782
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  37.32155     3.05464   12.218 5.84e-13 ***
## factor(am)1  -0.02362     1.54565    -0.015  0.988
## wt           -5.35281     0.78824    -6.791 1.87e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.098 on 29 degrees of freedom
## Multiple R-squared:  0.7528, Adjusted R-squared:  0.7358
## F-statistic: 44.17 on 2 and 29 DF,  p-value: 1.579e-09
```

We can see in this case, am is not as important as weight.