Regression Course Project

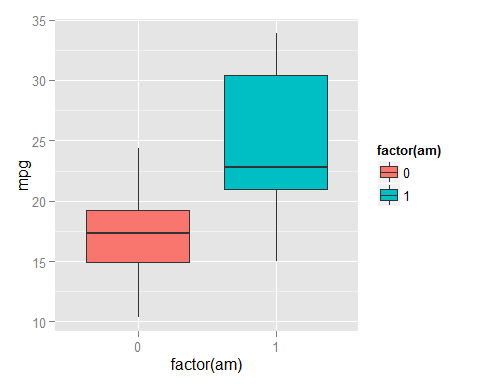
### SYNOPSIS

This is an analysis of mtcars dataset from 1974 Motor Trend US magazine. This dataset contains fues consumption (mpg) and 10 other variables for 32 different automobiles. We explore the relationship between the explanatory variables and the response variable (mpg) using the linear regression techniques that we learnet in the Regression class (From the Data Science track) in Coursera. Based on our findings, we attempt to answer the following questions:

* "Is an automatic or manual transmission better for MPG"
* "Quantify the MPG difference between automatic and manual transmissions"

### EXPLORATORY ANALYSIS OF DATA

We start by exploring the relationship between mpg and am (automatic/manual). From the plot, we can observe that the mpg for manual automobiles is larger than that of the automatic automobiles. However, there are many other variables that might impact the relationship.



### MODEL SELECTION

Next we will perform a simple linear regression using mpg as response variable, and all remaining attributes as explanatory variable. The objective here is to understand the linear relationship between reponse and explanatory variables, and create a baseline against which all following exclusions/additions will be compared.

We will fit a linear regression model using following R code **lm(mpg~.,data=mtcars)**.

When we look at the coefficients in Appendix 1 We can see that none of coefficient of the variables are significant.

Also the residual standard error of the fit is 2.650197, which we will use as baseline. In the subsequent steps, we will attempt to reduce this number:

Next, we perform stepwise regression (Appendix 2) to get a better indication of which variables seem to contribute to the changes mpg, and have significant p-value and high AIC (Area Under Curve).

## [1] "mpg ~ wt + qsec + am"

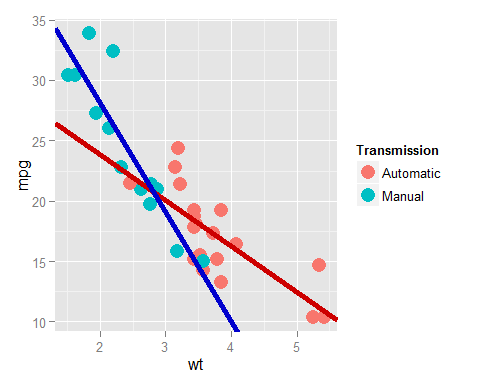
Here is the new list of coefficient based on the regression on the model recommended using stepwise procedure. As expected, they all seem to have significant p-values.

## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 9.617781 6.9595930 1.381946 1.779152e-01  
## wt -3.916504 0.7112016 -5.506882 6.952711e-06  
## qsec 1.225886 0.2886696 4.246676 2.161737e-04  
## am 2.935837 1.4109045 2.080819 4.671551e-02

Also, the residual has improved (decreased) to 2.4588465

In an attempt to further improve the model, we check for possible interaction relationships between the those 3 explanatory variables (am, wt and qsec). We know that mpg of an automobile is inversely related to the weight of the variable. Therefore, in the next plot, we check whether the relationship between mpg and weight is different for automatic and manual transmission, i.e., as weight increase, the mpg decreases at different rates for automatic and manual transmission. (Here we fit the variables weight and am agaist mpg)

## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 31.416055 3.0201093 10.402291 4.001043e-11  
## wt -3.785908 0.7856478 -4.818836 4.551182e-05  
## factor(am)1 14.878423 4.2640422 3.489277 1.621034e-03  
## wt:factor(am)1 -5.298360 1.4446993 -3.667449 1.017148e-03



The red curve is the fit for automatic transmission and the intercept is 31.42, and the slope is -3.78. The blue curve is fit for manual transmission, and we notice that the intercept increases by 14.8 units. Also, the slope decreases by -5.3 over that of automatic. This means that the rate of change of mpg in automatic/manual transmission is different for different weights. Give example: \*\*\*\*\*\*\*\*\*\* Automatic vs manual using different weights

## APPENDIX

### 1. Regression using all varibles

## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 12.30337416 18.71788443 0.6573058 0.51812440  
## cyl -0.11144048 1.04502336 -0.1066392 0.91608738  
## disp 0.01333524 0.01785750 0.7467585 0.46348865  
## hp -0.02148212 0.02176858 -0.9868407 0.33495531  
## drat 0.78711097 1.63537307 0.4813036 0.63527790  
## wt -3.71530393 1.89441430 -1.9611887 0.06325215  
## qsec 0.82104075 0.73084480 1.1234133 0.27394127  
## vs 0.31776281 2.10450861 0.1509915 0.88142347  
## am 2.52022689 2.05665055 1.2254035 0.23398971  
## gear 0.65541302 1.49325996 0.4389142 0.66520643  
## carb -0.19941925 0.82875250 -0.2406258 0.81217871

### 2. Stepwise regression

## Stepwise Model Path   
## Analysis of Deviance Table  
##   
## Initial Model:  
## mpg ~ cyl + disp + hp + drat + wt + qsec + vs + am + gear + carb  
##   
## Final Model:  
## mpg ~ wt + qsec + am  
##   
##   
## Step Df Deviance Resid. Df Resid. Dev AIC  
## 1 21 147.4944 70.89774  
## 2 - cyl 1 0.07987121 22 147.5743 68.91507  
## 3 - vs 1 0.26852280 23 147.8428 66.97324  
## 4 - carb 1 0.68546077 24 148.5283 65.12126  
## 5 - gear 1 1.56497053 25 150.0933 63.45667  
## 6 - drat 1 3.34455117 26 153.4378 62.16190  
## 7 - disp 1 6.62865369 27 160.0665 61.51530  
## 8 - hp 1 9.21946935 28 169.2859 61.30730