Regression Models

Jill Beck

April 3, 2016

Executive Summary

You work for Motor Trend, a magazine about the automobile industry. Looking at a data set of a collection of cars, they are interested in exploring the relationship between a set of variables and miles per gallon (MPG) (outcome). They are particularly interested in the following two questions:

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

Based on the final model fitting results, we can conclude that:

1. As wt increases per 1000lb (0.5 tons), MPG decreases by 2.5.
2. MPG will decrease very slighly with an increase in horsepower (HP).
3. 'cyl' increases from 4 to 6 to 8. This will cause MPG to decrease respectively by 3 and 2 times.
4. Automatic gearing has higher MPG when compared to manual gearing.

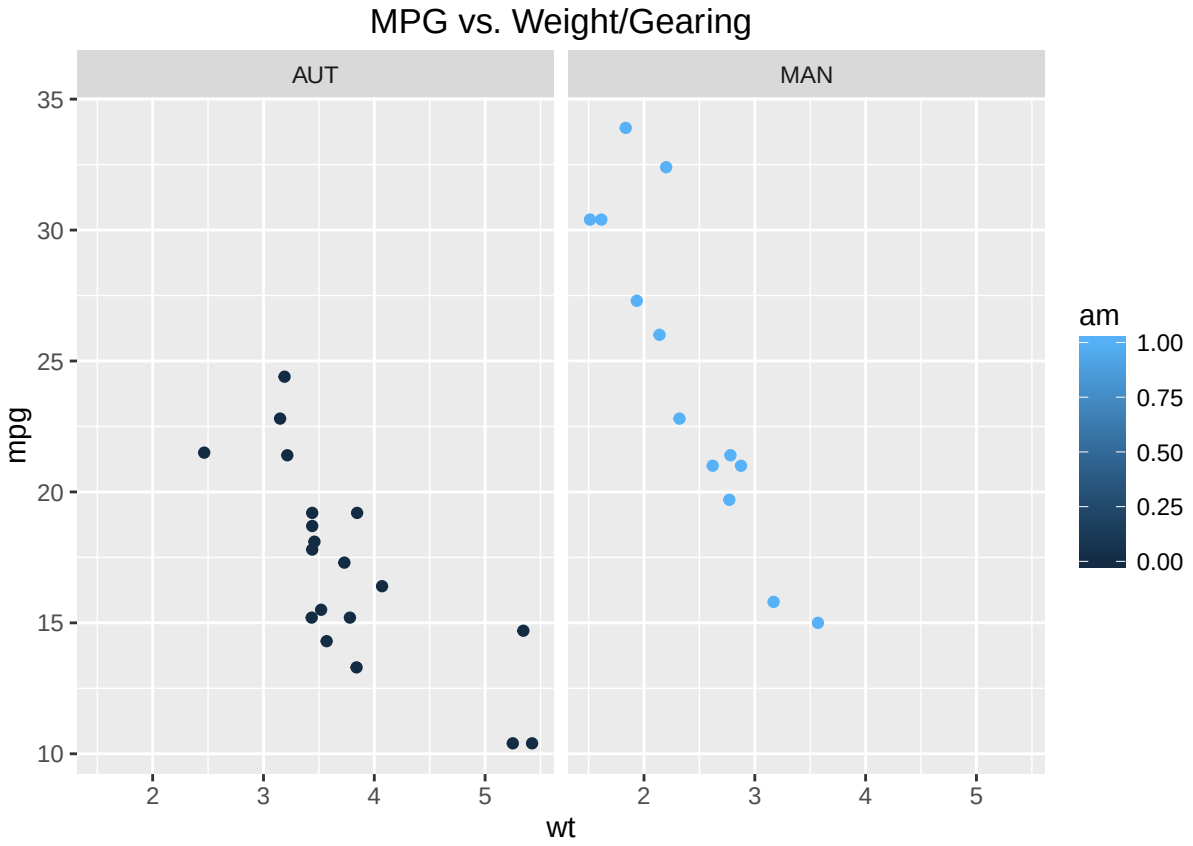
Data Analysis

```
library(ggplot2)
data(mtcars)
attach(mtcars)
```

```
## The following object is masked from package:ggplot2:
##
##     mpg
```

This first look at the data shows us that horsepower and MPG are inversely related and manual transmission cars are generally more fuel efficient.

```
mtcars$amf[am==0]='AUT'
mtcars$amf[am==1]='MAN'
print(qplot(x=wt, y=mpg, colour=am, facets=.~amf, data=mtcars,main="MPG vs. Weight/Gearing"))
```



The initial observation is that cars with a manual transmission tend to use a smaller amount of fuel and weigh less.

Creating The Model

We will perform linear regression with all variables, and then perfom stepwise model selection to select best predictors. The final model will incorporate 'cyl', 'wt', 'hp' and 'am'.

```
fit<-glm(mpg~as.factor(cyl) + as.factor(vs) + as.factor(am) + as.factor(gear) + as.factor(carb) + disp +
library(MASS)
step <- stepAIC(fit, direction="both")
```

```
## Start:  AIC=169.22
## mpg ~ as.factor(cyl) + as.factor(vs) + as.factor(am) + as.factor(gear) +
##     as.factor(carb) + disp + hp + drat + wt + qsec
##
##                    Df Deviance    AIC
## - as.factor(carb)  5   134.00 162.64
## - as.factor(gear)  2   124.38 166.25
## - as.factor(am)    1   121.55 167.52
## - qsec             1   121.64 167.54
## - drat             1   122.22 167.70
## - as.factor(cyl)   2   131.33 168.00
## - as.factor(vs)    1   124.03 168.17
## <none>                 120.40 169.22
## - disp             1   130.37 169.76
```

```
## - wt               1   145.96 173.37
## - hp               1   146.07 173.40
##
## Step:  AIC=162.64
## mpg ~ as.factor(cyl) + as.factor(vs) + as.factor(am) + as.factor(gear) +
##     disp + hp + drat + wt + qsec
##
##                   Df Deviance    AIC
## - as.factor(gear)  2   139.02 159.82
## - disp             1   135.00 160.88
## - drat             1   135.19 160.92
## - as.factor(vs)    1   137.68 161.51
## - as.factor(cyl)   2   146.57 161.51
## - qsec             1   139.26 161.87
## <none>                 134.00 162.64
## - as.factor(am)    1   145.93 163.37
## - wt               1   153.80 165.05
## - hp               1   156.79 165.67
## + as.factor(carb)  5   120.40 169.22
##
## Step:  AIC=159.82
## mpg ~ as.factor(cyl) + as.factor(vs) + as.factor(am) + disp +
##     hp + drat + wt + qsec
##
##                   Df Deviance    AIC
## - drat             1   139.99 158.04
## - as.factor(cyl)   2   149.45 158.13
## - disp             1   140.57 158.17
## - as.factor(vs)    1   141.21 158.32
## - qsec             1   142.66 158.64
## <none>                 139.02 159.82
## - as.factor(am)    1   155.59 161.42
## - hp               1   157.20 161.75
## + as.factor(gear)  2   134.00 162.64
## - wt               1   170.21 164.29
## + as.factor(carb)  5   124.38 166.25
##
## Step:  AIC=158.04
## mpg ~ as.factor(cyl) + as.factor(vs) + as.factor(am) + disp +
##     hp + wt + qsec
##
##                   Df Deviance    AIC
## - disp             1   141.24 156.32
## - as.factor(vs)    1   142.33 156.57
## - as.factor(cyl)   2   152.32 156.74
## - qsec             1   143.09 156.74
## <none>                 139.99 158.04
## + drat             1   139.02 159.82
## - hp               1   157.73 159.86
## - as.factor(am)    1   159.46 160.21
## + as.factor(gear)  2   135.19 160.92
## - wt               1   170.71 162.39
## + as.factor(carb)  5   126.94 164.91
##
```

```
## Step:  AIC=156.32
## mpg ~ as.factor(cyl) + as.factor(vs) + as.factor(am) + hp + wt +
##     qsec
##
##                   Df Deviance    AIC
## - qsec             1   143.68 154.87
## - as.factor(vs)    1   143.98 154.94
## - as.factor(cyl)   2   159.82 156.28
## <none>                 141.24 156.32
## + disp             1   139.99 158.04
## + drat             1   140.57 158.17
## - hp               1   159.42 158.20
## - as.factor(am)    1   160.12 158.34
## + as.factor(gear)  2   136.55 159.24
## - wt               1   180.88 162.24
## + as.factor(carb)  5   138.91 165.79
##
## Step:  AIC=154.87
## mpg ~ as.factor(cyl) + as.factor(vs) + as.factor(am) + hp + wt
##
##                   Df Deviance    AIC
## - as.factor(vs)    1   151.03 154.47
## <none>                 143.68 154.87
## - as.factor(cyl)   2   168.96 156.06
## + qsec             1   141.24 156.32
## - as.factor(am)    1   160.12 156.34
## + disp             1   143.09 156.74
## + drat             1   143.35 156.80
## + as.factor(gear)  2   140.24 158.10
## - hp               1   180.02 160.09
## - wt               1   184.77 160.92
## + as.factor(carb)  5   140.20 164.09
##
## Step:  AIC=154.47
## mpg ~ as.factor(cyl) + as.factor(am) + hp + wt
##
##                   Df Deviance    AIC
## <none>                 151.03 154.47
## - as.factor(am)    1   160.78 154.47
## + as.factor(vs)    1   143.68 154.87
## + qsec             1   143.98 154.94
## - as.factor(cyl)   2   180.29 156.13
## + disp             1   150.41 156.34
## + drat             1   150.81 156.42
## + as.factor(gear)  2   149.66 158.18
## - hp               1   182.97 158.61
## - wt               1   197.20 161.00
## + as.factor(carb)  5   145.39 163.25
```

```
step$anova
```

```
## Stepwise Model Path
## Analysis of Deviance Table
##
```

```
## Initial Model:
## mpg ~ as.factor(cyl) + as.factor(vs) + as.factor(am) + as.factor(gear) +
##     as.factor(carb) + disp + hp + drat + wt + qsec
##
## Final Model:
## mpg ~ as.factor(cyl) + as.factor(am) + hp + wt
##
##
##                 Step Df   Deviance Resid. Df Resid. Dev      AIC
## 1                                         15   120.4027 169.2155
## 2 - as.factor(carb)  5 13.5988573        20   134.0015 162.6398
## 3 - as.factor(gear)  2  5.0215145        22   139.0230 159.8170
## 4            - drat  1  0.9672159        23   139.9903 158.0388
## 5            - disp  1  1.2473996        24   141.2377 156.3227
## 6            - qsec  1  2.4420033        25   143.6797 154.8713
## 7   - as.factor(vs)  1  7.3459298        26   151.0256 154.4669
```

Conclusions

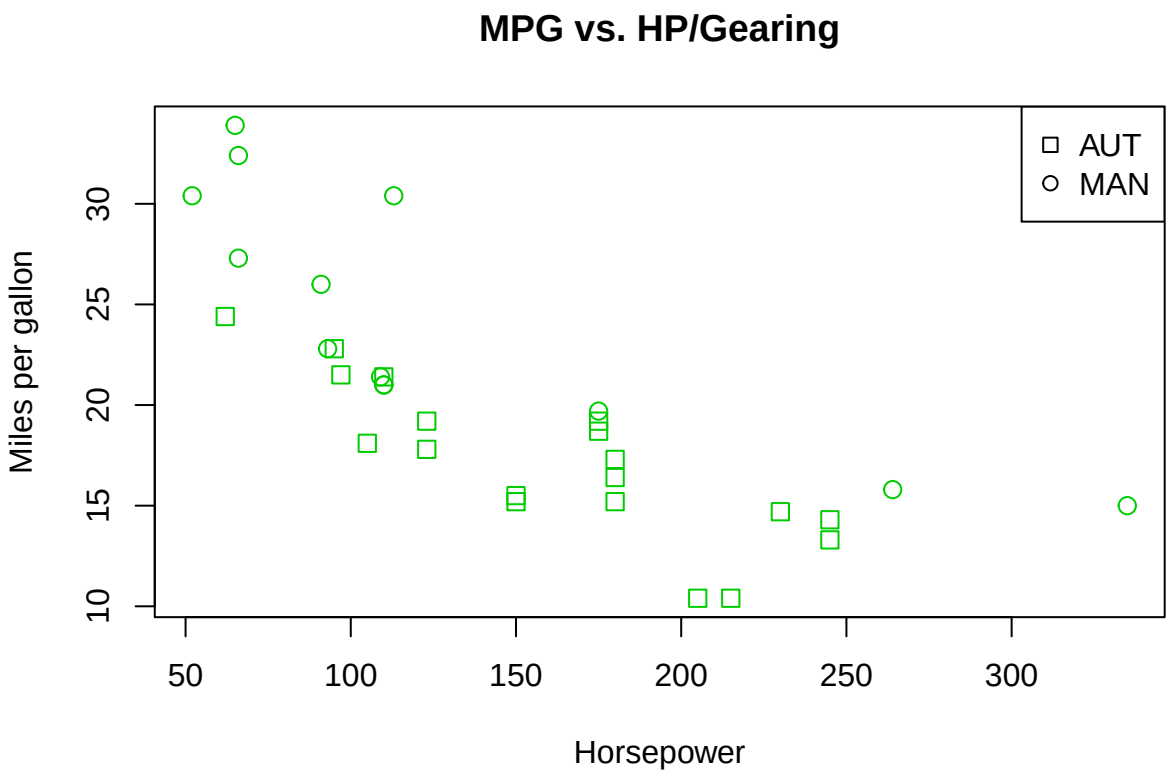
Based on the final results, we can conclude that:

1. As wt increases per 1000lb (0.5 tons), MPG decreases by 2.5.
2. MPG will decrease very slighly with an increase in horsepower.
3. 'cyl' increases from 4 to 6 to 8. This will cause MPG to decrease respectively by 3 and 2 times.
4. Automatic gearing has higher MPG when compared to manual gearing.

Appendix

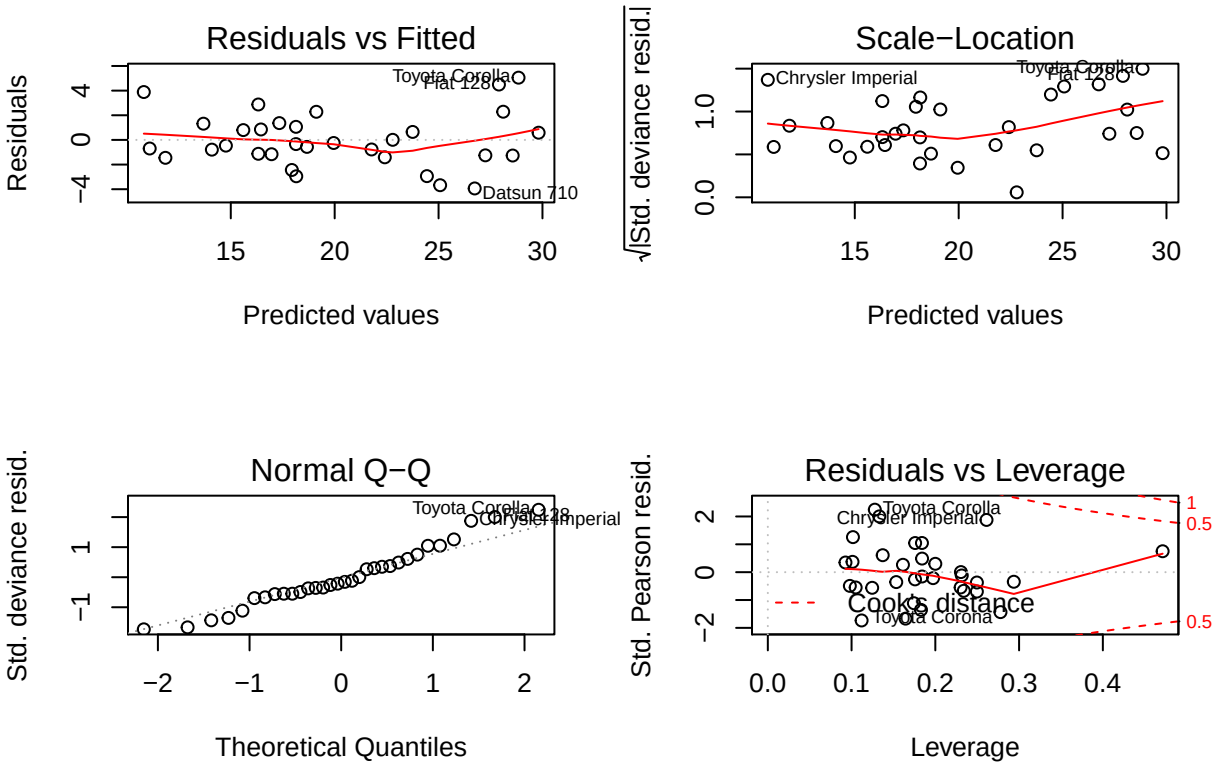
Initial Explorations

```
plot(hp, mpg, pch=am,col=259,bg=7,
  xlab="Horsepower", cex=1.2,
  ylab="Miles per gallon", main="MPG vs. HP/Gearing")
legend("topright", c("AUT","MAN"), pch=c(0,1))
```



Final Model

```
fit2<-glm(mpg ~ as.factor(cyl) + as.factor(am) + hp + wt, data=mtcars)
layout(matrix(c(1,2,3,4),2,2))
plot(fit2)
```


Residuals vs Fitted
Residuals
Predicted values
Toyota Corolla
Fiat 128
Datsun 710
Scale−Location
√|Std. deviance resid.|
Chrysler Imperial
Normal Q−Q
Std. deviance resid.
Theoretical Quantiles
Residuals vs Leverage
Std. Pearson resid.
Cook's distance
Leverage
Toyota Corona