# Regression Models Project

*Robert Tuck*

*June 21, 2015*

### Background

### Data processing

```
car_data <- data.frame(mtcars)
```

### Data Exploration

```
car_data$cyl <- factor(car_data$cyl)
car_data$vs <- factor(car_data$vs)
car_data$gear <- factor(car_data$gear)
car_data$carb <- factor(car_data$carb)
car_data$am <- factor(car_data$am,labels=c('Automatic','Manual'))
```

### Regression Model

Perform a regression on mpg as an outcome of all the predictors and then step through the regressions to look for the best model.

```
initialmodel <- lm(mpg ~ ., data = car_data)
bestmodel <- step(initialmodel, direction = "both")
```

The adjusted R-squared value of 0.84 which is the maximum obtained considering all combinations of variables. From these results we can conclude that more than 84% of the variability is explained by the above model.

### Inference

Perform a t-test to test for transmission type vs mpg.

```
t.test(mpg ~ am, data = car_data)
```

```
##
##  Welch Two Sample t-test
##
## data:  mpg by am
## t = -3.7671, df = 18.332, p-value = 0.001374
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.280194  -3.209684
## sample estimates:
## mean in group Automatic    mean in group Manual
##                17.14737                24.39231
```

Based on the t-test results, we reject the null hypothesis that the mpg distributions for manual and automatic transmissions are the same. #### Conclusion

Based on the analysis done in this project, we can conclude that:

- Cars with Manual transmission get 1.8 more miles per gallon compared to cars with Automatic transmission. (1.8 adjusted for hp, cyl, and wt).
- mpg will decrease by 2.5 for every 1000 lb increase in wt.
- mpg decreases negligibly (only 0.32) with every increase of 10 in hp.
- If number of cylinders, cyl increases from 4 to 6 and 8, mpg will decrease by a factor of 3 and 2.2 respectively (adjusted by hp, wt, and am).

### Appendix

### Data Exploration

```
summary(car_data)
```

```
##       mpg        cyl         disp             hp             drat
##  Min.   :10.40   4:11   Min.   : 71.1   Min.   : 52.0   Min.   :2.760
##  1st Qu.:15.43   6: 7   1st Qu.:120.8   1st Qu.: 96.5   1st Qu.:3.080
##  Median :19.20   8:14   Median :196.3   Median :123.0   Median :3.695
##  Mean   :20.09          Mean   :230.7   Mean   :146.7   Mean   :3.597
##  3rd Qu.:22.80          3rd Qu.:326.0   3rd Qu.:180.0   3rd Qu.:3.920
##  Max.   :33.90          Max.   :472.0   Max.   :335.0   Max.   :4.930
##        wt             qsec       vs             am     gear   carb
##  Min.   :1.513   Min.   :14.50   0:18   Automatic:19   3:15   1: 7
##  1st Qu.:2.581   1st Qu.:16.89   1:14   Manual   :13   4:12   2:10
##  Median :3.325   Median :17.71                         5: 5   3: 3
##  Mean   :3.217   Mean   :17.85                                4:10
##  3rd Qu.:3.610   3rd Qu.:18.90                                6: 1
##  Max.   :5.424   Max.   :22.90                                8: 1
```

### Regression Model

Summary of bestmodel data

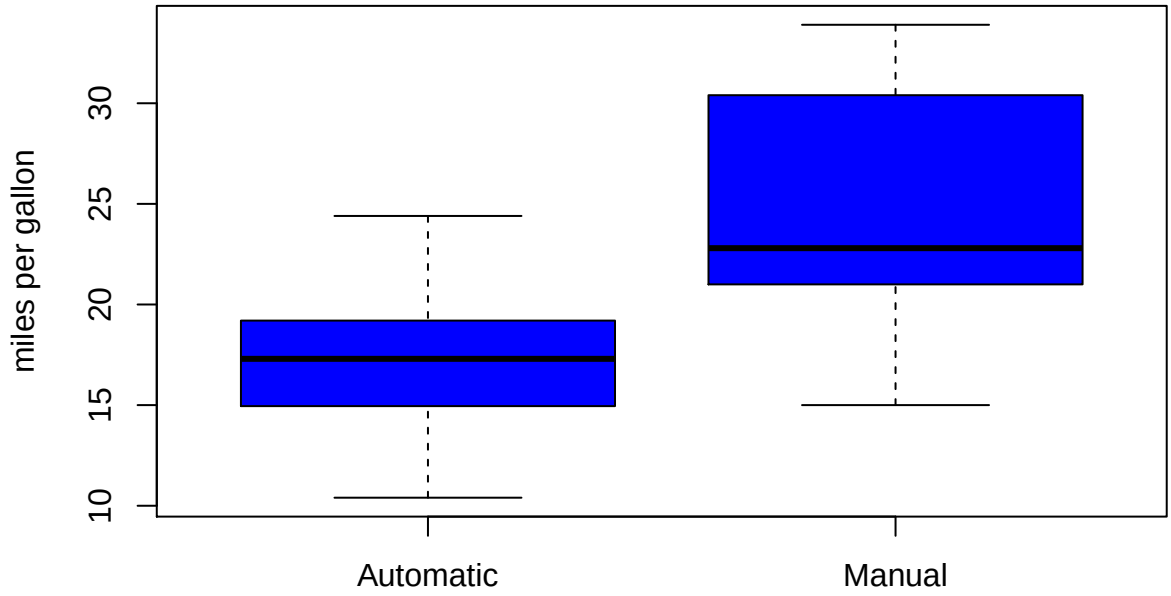
```
summary(bestmodel)
```

```
##
## Call:
## lm(formula = mpg ~ cyl + hp + wt + am, data = car_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.9387 -1.2560 -0.4013  1.1253  5.0513
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 33.70832    2.60489  12.940 7.73e-13 ***
```

```
## cyl6        -3.03134    1.40728  -2.154  0.04068 *
## cyl8        -2.16368    2.28425  -0.947  0.35225
## hp          -0.03211    0.01369  -2.345  0.02693 *
## wt          -2.49683    0.88559  -2.819  0.00908 **
## amManual     1.80921    1.39630   1.296  0.20646
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.41 on 26 degrees of freedom
## Multiple R-squared:  0.8659, Adjusted R-squared:  0.8401
## F-statistic: 33.57 on 5 and 26 DF,  p-value: 1.506e-10
```

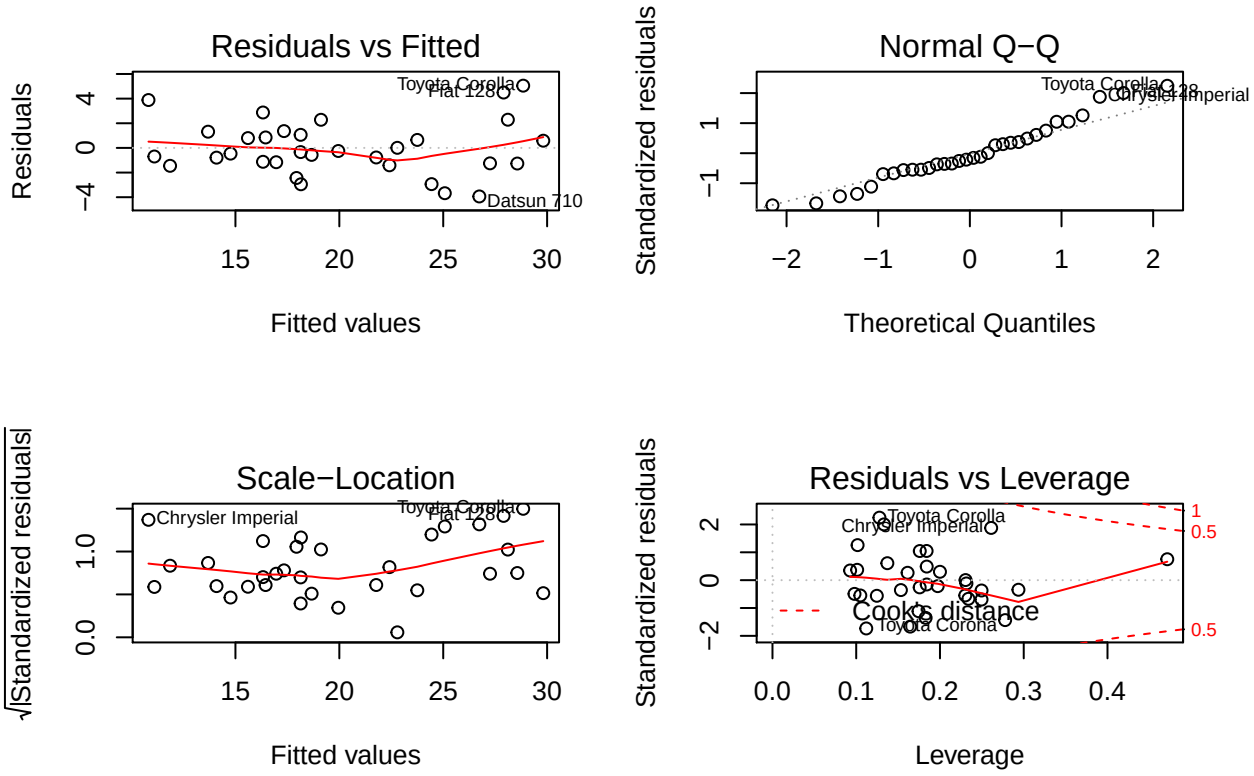
Boxplot showing relationships between transmission types

```
boxplot(mpg ~ am, data = car_data, col = "blue", ylab = "miles per gallon")
```




Residual plots of the the bestmodel

```
par(mfrow=c(2, 2))
plot(bestmodel)
```


Residuals vs Fitted
Residuals
Fitted values
Toyota Corolla
Fiat 128
Datsun 710
Normal Q−Q
Standardized residuals
Theoretical Quantiles
Toyota Corolla
Chrysler Imperial
Scale−Location
√|Standardized residuals|
Fitted values
Chrysler Imperial
Toyota Corolla
Fiat 128
Residuals vs Leverage
Standardized residuals
Leverage
Cook's distance
Toyota Corolla
Chrysler Imperial
Toyota Corona