Regression Project

Context

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"

"Quantifying how different is the MPG between automatic and manual transmissions?"

Additional Information

Take the mtcars data set and write up an analysis to answer their question using regression models and exploratory data analyses.

Your report must be:

Written as a PDF printout of a compiled (using knitr) R markdown document. Do not use any packages that are not in R-base or the library datasets. Brief. Roughly the equivalent of 2 pages or less for the main text. Supporting figures in an appendix can be included up to 5 total pages including the 2 for the main report. The appendix can only include figures. Include a first paragraph executive summary.

Executive Summary

This week Motor Trend is going to look at the affect of automatic transmissions on fuel efficiency. To do this we will use a data set that examines the fuel efficency and 10 aspects of automobile design and performance for 32 automobiles (all 1973 - 1974 models). Out of the 32 cars, 13 have manual transmissions and 19 have automatic transmissions.

In this data set on average there is a difference in fuel efficency depending on transmission type such that on average manual vehicles achieve a fuel effiency of 7.2 miles per gallon more than automatic vehicles.

However, transmission type is not a particularly good predictor of fuel efficiency. If we add transmission type to this model, then the difference in fuel effiency for a manual transmission is much smaller, just 0.18 miles per gallon for a vehicle with the same weight and number of cylinders.

Therefore we conclude that number of cylinders and weight are good predictors of fuel efficiency, but transmission type is not.

load the mtcars dataset

```
data(mtcars)

#See appendix for exploratory analysis.

summary(mtcars)
```

From the summary results I can correctly parse the data to create the regression model.

```
mtcars$drat <- factor(mtcars$drat)

mtcars$cyl <- factor(mtcars$cyl)

mtcars$vs <- factor(mtcars$vs)

mtcars$gear <- factor(mtcars$gear)
```

```
mtcars$carb <- factor(mtcars$carb)

mtcars$am <- factor(mtcars$am,labels=c('Automatic','Manual'))
```

REGRESSION MODEL

Compare the full model to

```
#Result shown in the Appendix

full.model <- lm(mpg ~ ., data = mtcars)

best.model <- step(full.model, direction = "backward")

#Result shown in the Appendix

summary(best.model)
```

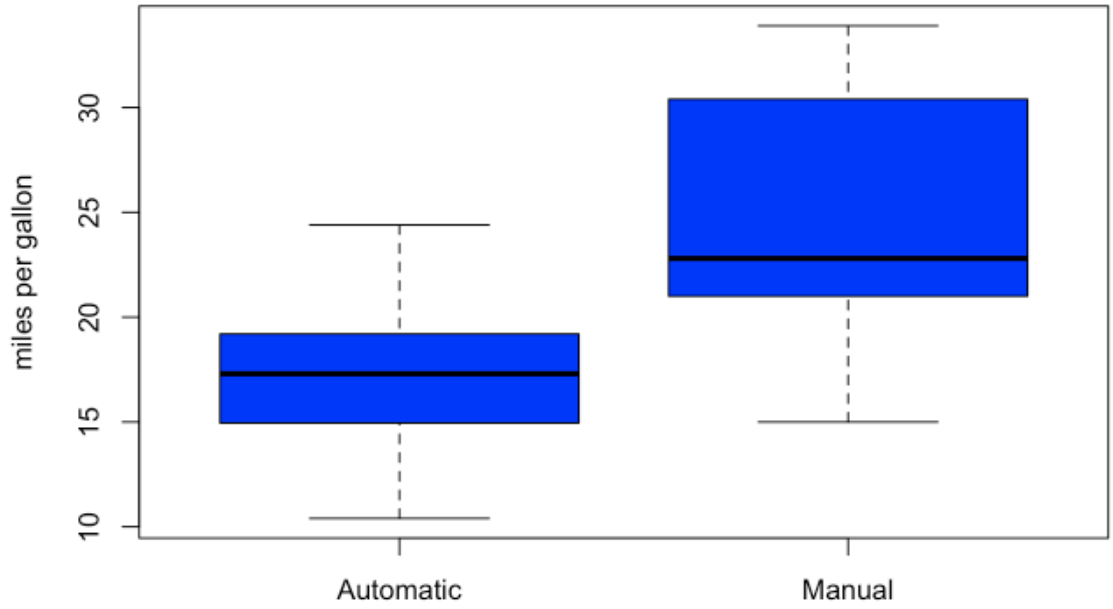
This procedure determines that the best model includes the cyl6, cyl8, hp, wt, and amManual variables (overall p-value<0.001). The adjusted R-squared indicates that about 84% of the variance is explained by the final model. Moreover, the output of this model suggests that mpg decreases with respect to cylinders (-3.03 and -2.16 for cyl6 and cyl8, respectively), horsepower (-0.03), and weight (for every 1,000lb, by -2.5). On the other hand, mpg increases with respect to having a manual transmission (by 1.8). Residual plots (see appendix) suggest that some transformation may be necessary to achieve linearity.

A test is then run to determine the difference in mpg between the automatic and manual transmissions.

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

```
##

##  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
```

```
#Result shown in the Appendix

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



The boxplot confirms the result of the t-test =, as the difference in MPG between the automatic and manuale transmissions are significatnly different (p-value < 0.05).

Conclusion

Although in this data set on average manual vehicles achieve a fuel effiency of 7.2 miles per gallon more than automatic vehicles, transmission type is not a particularly good predictor of fuel efficiency. We were able to identify that the number of cylinders and the weight of the automobile are good predictors of fuel efficiency, achieving an adjusted R squared of 0.82. If we add transmission type to this model, then the difference in fuel effiency for a manual transmission is much smaller, just 0.18 miles per gallon for a vehicle with the same weight and number of cylinders. Therefore we conclude that number of cylinders and displacement are good predictors of fuel efficiency, but transmission type is not.

Appendix

Explorartoty Analysis

```
summary(mtcars)
```

```
##       mpg        cyl         disp             hp             drat
##  Min.   :10.40   4:11   Min.   : 71.1   Min.   : 52.0   3.07   : 3
##  1st Qu.:15.43   6: 7   1st Qu.:120.8   1st Qu.: 96.5   3.92   : 3
##  Median :19.20   8:14   Median :196.3   Median :123.0   2.76   : 2
##  Mean   :20.09          Mean   :230.7   Mean   :146.7   3.08   : 2
##  3rd Qu.:22.80          3rd Qu.:326.0   3rd Qu.:180.0   3.15   : 2
##  Max.   :33.90          Max.   :472.0   Max.   :335.0   3.9    : 2
##                                                         (Other):18
##        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 Results

```
summary(best.model)
```

```
##
## Call:
## lm(formula = mpg ~ cyl + disp + hp + drat + wt + gear, data = mtcars)
##
## Residuals:
##           Mazda RX4       Mazda RX4 Wag          Datsun 710
##           6.305e-01          -6.305e-01           2.776e-17
##      Hornet 4 Drive   Hornet Sportabout             Valiant
##          -1.205e-01          -9.637e-02           1.205e-01
##          Duster 360           Merc 240D            Merc 230
##           1.665e-16          -3.053e-16           1.305e-15
##            Merc 280           Merc 280C          Merc 450SE
##           7.000e-01          -7.000e-01          -9.385e-01
##          Merc 450SL         Merc 450SLC  Cadillac Fleetwood
##           1.643e+00          -7.044e-01           2.776e-17
## Lincoln Continental   Chrysler Imperial            Fiat 128
##          -5.551e-17           0.000e+00           1.686e+00
##         Honda Civic      Toyota Corolla       Toyota Corona
##          -3.608e-16          -4.718e-16          -1.943e-16
##    Dodge Challenger         AMC Javelin          Camaro Z28
##          -1.205e-01           9.637e-02          -2.776e-16
##    Pontiac Firebird           Fiat X1-9       Porsche 914-2
```

```
##           1.205e-01          -1.686e+00           1.388e-16
##        Lotus Europa      Ford Pantera L        Ferrari Dino
##           0.000e+00           8.327e-17           0.000e+00
##       Maserati Bora          Volvo 142E
##           5.551e-17          -2.498e-16
##
## Coefficients: (1 not defined because of singularities)
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   31.038     20.528   1.512   0.2051
## cyl6         -59.745     40.381  -1.480   0.2131
## cyl8         -79.839     49.838  -1.602   0.1844
## disp          -1.389      1.072  -1.295   0.2649
## hp             3.259      2.443   1.334   0.2530
## drat2.93      20.944     27.242   0.769   0.4849
## drat3        -29.174     13.566  -2.151   0.0979 .
## drat3.07    -157.389    119.891  -1.313   0.2595
## drat3.08      34.305     25.051   1.369   0.2427
## drat3.15     -19.546     16.086  -1.215   0.2911
## drat3.21    -252.792    187.756  -1.346   0.2494
## drat3.23    -101.149     69.620  -1.453   0.2199
## drat3.54    -627.339    470.801  -1.332   0.2535
## drat3.62    -334.116    254.614  -1.312   0.2597
## drat3.69    -225.129    157.636  -1.428   0.2264
## drat3.7     -170.968    130.460  -1.311   0.2602
## drat3.73    -269.020    199.065  -1.351   0.2479
## drat3.77    -244.231    194.278  -1.257   0.2771
## drat3.85    -377.210    272.852  -1.382   0.2390
## drat3.9     -304.535    219.304  -1.389   0.2373
## drat3.92    -342.264    244.439  -1.400   0.2340
## drat4.08    -321.424    237.585  -1.353   0.2475
## drat4.11    -414.963    299.015  -1.388   0.2375
```

```
## drat4.22    -323.732    242.931  -1.333   0.2535
## drat4.43    -145.031    114.904  -1.262   0.2755
## drat4.93    -277.391    205.793  -1.348   0.2490
## wt             4.945      4.461   1.109   0.3298
## gear4        204.483    139.689   1.464   0.2171
## gear5             NA         NA      NA       NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.704 on 4 degrees of freedom
## Multiple R-squared:  0.9897, Adjusted R-squared:  0.9201
## F-statistic: 14.21 on 27 and 4 DF,  p-value: 0.009558
```

Boxplot

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