Regression Models Assignment

Mark C

Executive Summary

In this report, data gathered by Motor Trend US Magazine is explored and analyzed in order to determine the relationship between a set of variables and miles per gallon of automobiles. Here, we will try to answer the following two questions :

1) Is an automatic or manual transmission better for MPG?
2) What is the MPG difference between automatic and manual transmissions?

Summary of data

The data was extracted from the 1974 Motor Trend US magazine, and comprises fuel consumption and 10 aspects of automobile design and performance for 32 automobiles (1973-74 models). The qualitative variables such as number of cylinders and gears were converted to factors. To see a description of the variables , enter ?mtcars in the Rconsole.

Exploratory analysis on the data

A boxplot was produced to show the difference between automatic and manual in terms of MPG. In figure 1, it is clear that manual transmission produces more MPG. Next, a pairwise graph (figure 2) was created in order to get a greater intuition of what other variables may be of interest.

```
head(cov2cor(cov(sapply(mtcars, as.numeric))), 1)
```

```
##     mpg       cyl       disp         hp      drat         wt     qsec
## mpg   1 -0.852162 -0.8475514 -0.7761684 0.6811719 -0.8676594 0.418684
##            vs        am      gear       carb
## mpg 0.6640389 0.5998324 0.4802848 -0.6067431
```

There is a linear relationship between MPG and each of cyl, disp, hp, drat, wt, qsec, vs, am. The covariance was also computed above between every variable and the positive values were noted (qsec = 0.419, vs = 0.664, am = 0.600, gear = 0.480). Then a linear model was fit on all the variables to determine which variables should be used in the final models.

```
allAttr_model = lm(mpg ~ ., data = mtcars)
allAttr_model$coeff
```

```
## (Intercept)        cyl6        cyl8        disp          hp        drat
## 23.87913244 -2.64869528 -0.33616298  0.03554632 -0.07050683  1.18283018
##          wt        qsec         vs1    amManual       gear4       gear5
## -4.52977584  0.36784482  1.93085054  1.21211570  1.11435494  2.52839599
##       carb2       carb3       carb4       carb6       carb8
## -0.97935432  2.99963875  1.09142288  4.47756921  7.25041126
```

Above is the summary from this model is shown. The lowest p values were taken (i.e. wt = 0.063, am = 0.234, qsec = 0.274) due to their high significance in predicting MPG.

Model

From the initial model, covariance test and looking at the pairwise graph the following variables stood out in particular: qsec, vs, am, wt and gear. Next a stepwise modelling process was used in order to obtain the most significant predictors. This is done using the step function which creates regression models with different variables and produces list of the best predictors.

```
new_model <- step(lm(mpg ~ ., data = mtcars), trace = 0)
summary(new_model)$coef
```

```
##                Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 33.70832390 2.60488618 12.940421 7.733392e-13
## cyl6        -3.03134449 1.40728351 -2.154040 4.068272e-02
## cyl8        -2.16367532 2.28425172 -0.947214 3.522509e-01
## hp          -0.03210943 0.01369257 -2.345025 2.693461e-02
## wt          -2.49682942 0.88558779 -2.819404 9.081408e-03
## amManual     1.80921138 1.39630450  1.295714 2.064597e-01
```

As shown above, the most significant predictors in determining the MPG are cyl, hp, wt and am.

```
new_model <- step(lm(mpg ~ ., data = mtcars), trace = 0)
new_model$coeff
```

```
## (Intercept)        cyl6        cyl8          hp          wt    amManual
## 33.70832390 -3.03134449 -2.16367532 -0.03210943 -2.49682942  1.80921138
```

The summary for this model is show above, in particular the forumla is given as: lm(formula = mpg ~ cyl + hp + wt + am, data = mtcars). This selection model yielded an R squared value of 84%, meaning that very high percentage of variation is explained by the regression model. Next, the new model was compared with a basic model that only uses transmission type as its predictor.

```
basic_model <- lm(mpg ~ am, data = mtcars)
compare <- anova(basic_model, new_model)
compare$Pr
```

```
## [1]           NA 1.688435e-08
```

A p-value of 1.688e-08 was obtained (figure 3). This value is miniscule which means that the added predictors are significant to improving the model's accuracy.

Diagnostics

The residuals from the final model are plotted in Figure 3. - The Residuals vs Fitted plot shows no pattern between the residuals and fitted values indicating that this regression model is well fit.The QQ plot shows that the points line up as expected meaning that the distribtion is normal and our model predictions are accurate.In both the Scale-Location plot and the Residuals vs Leverage plots, the points are in a group with none too far from the center indicating no point had too much leverage.

Statistical Inference

A two sample t-test was conducted between the different transmission types. The null hypothesis is that the transmission types do have an effect on the MPG.

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

```
##
##  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
```

The results are shown above. The low p-value and difference in means show that manual transmission has significantly more MPG than automatic.

Conclusions

The transmission type of a car has a significant effect on its fuel efficiency. According to the model, manual transmission, on average, has 1.81 MPG more than automatic. According to the boxplot, manual transmission produces more MPG than automatic transmission.

Apendix

Figure 1

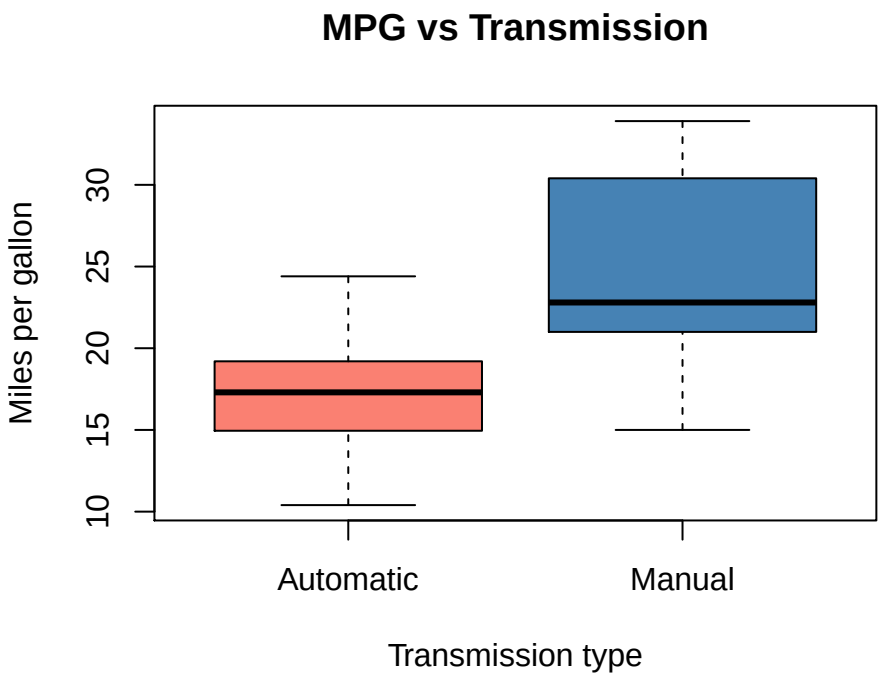


Figure 2

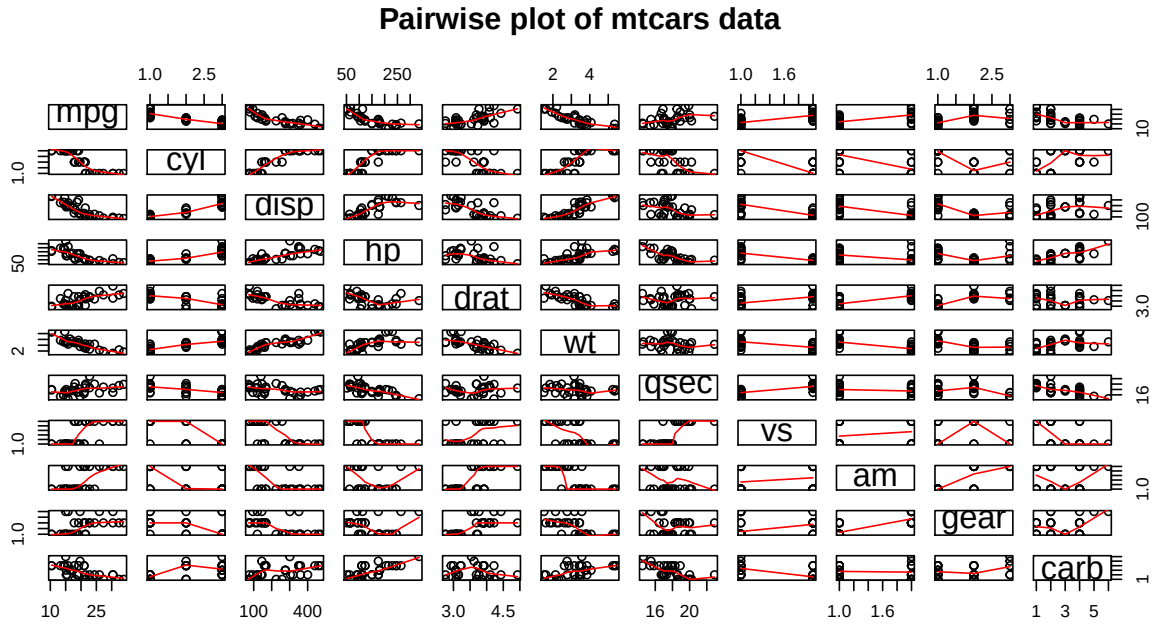
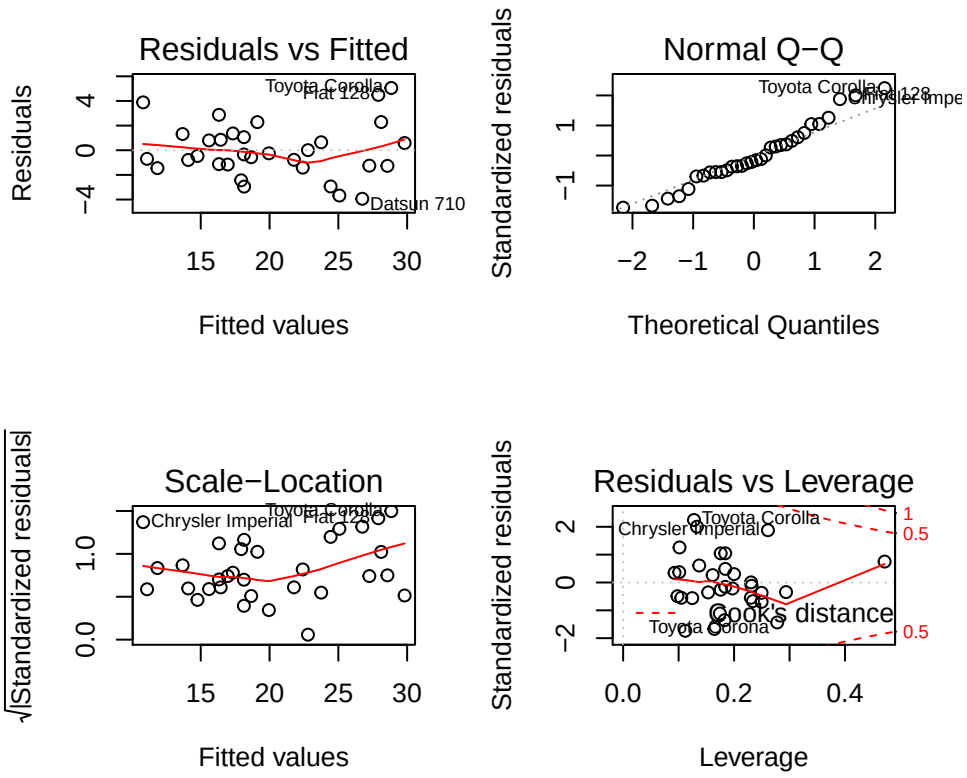


Figure 3


Residuals vs Fitted
Residuals
Fitted values
Toyota Corolla
Fiat 128
Datsun 710
Normal Q−Q
Standardized residuals
Theoretical Quantiles
Toyota Corolla
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