Regression Models Class Project: Automatic vs Manual Transmission

Joselle Abagat

7/8/2018

Topic:

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”

Analysis and Summary

Loaded Libraries

Executive Summary

This project explores the mtcars dataset in order to determine whether automatic or manual transmission is better for MPG. A quick data summary shows that manual is transmission is better for MPG. However, this is strictly looking at MPG as a function of transmission and nothing more. It does not take into account other variables such as cylinders, horsepower, weight, etc. This project will fit and select different models, perform residual plotting and diagnostics, and perform inference in an attempt to quantify the MPG difference between automatic and manual transmissions.

Explore mtcars data set

The data frame mtcars consists of 32 observations on the following 11 variables: mpg, cyl, disp, hp, drat, wt, qsec, vs, am, gear, carb.

```
str(mtcars)
```

```
## Classes 'data.table' and 'data.frame':   32 obs. of  11 variables:
##  $ mpg : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
##  $ cyl : num  6 6 4 6 8 6 8 4 4 6 ...
##  $ disp: num  160 160 108 258 360 ...
##  $ hp  : num  110 110 93 110 175 105 245 62 95 123 ...
##  $ drat: num  3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...
##  $ wt  : num  2.62 2.88 2.32 3.21 3.44 ...
##  $ qsec: num  16.5 17 18.6 19.4 17 ...
##  $ vs  : num  0 0 1 1 0 1 0 1 1 1 ...
##  $ am  : num  1 1 1 0 0 0 0 0 0 0 ...
##  $ gear: num  4 4 4 3 3 3 3 4 4 4 ...
##  $ carb: num  4 4 1 1 2 1 4 2 2 4 ...
##  - attr(*, ".internal.selfref")=<externalptr>
```

We want to explore the relationship between mpg and am (Transmission: 0 = automatic, 1 = manual). Based on the data summaries in Appendix A, mpg ranges from [10.4 to 24.4] with an average of 17.15 for automatic

transmission while mpg ranges from [15 to 33.9] with an average of 24.39 for manual transmission. This can allow us to conclude that driving a manual transmission car leads to more miles per gallon. However, this does not take into account the effect of other variables such as the number of cylinders or car weight.

```
cor(mtcars$mpg, mtcars[, -1])
```

```
##            cyl       disp         hp      drat         wt     qsec
## [1,] -0.852162 -0.8475514 -0.7761684 0.6811719 -0.8676594 0.418684
##             vs        am      gear       carb
## [1,] 0.6640389 0.5998324 0.4802848 -0.5509251
```

The correlation shows us that while mpg is positively impacted by transmission and number of gears, it is inversely impacted by the number of cylinders, horsepower, weight.

Model Fitting and Selection

Analyzing different model fits, a general linear model did not apply since mpg can be represented as arange of values. Therefore, a binomial linear model could not be applied. It is also not a rate, therefore, a poisson linear model should not be applied. The most appropriate is a simple linear model via function lm(). Using step(), we can use a stepwise algorithm to choose the best model by AIC.

```
mdlAll <- lm(mpg ~ ., mtcars)
mdlBest <- step(mdlAll, trace = 0)
```

Looking at the summaries of mdlAll and mdlBest in Appendix B, it can be seen that using all the predictors result in P-values > 0.05, which show that the relationships are not significant. Using the stepwise algorithm, the best model shows that wt (weight), qsec (1/4 mile time), and am (transmission) are the relevant predictors.

Diagnostics

In order to test whethere the predictors are truly significant, let's compute the analysis of variance on the fitted model, mdlBest. The Analysis of Variance Table below shows that weight, quarter mile time and transmission are significant in response to mpg (all p-values are << 0.05).

```
anova(mdlBest)
```

```
## Analysis of Variance Table
##
## Response: mpg
##           Df Sum Sq Mean Sq  F value    Pr(>F)
## wt         1 847.73  847.73 140.2143 2.038e-12 ***
## qsec       1  82.86   82.86  13.7048 0.0009286 ***
## am         1  26.18   26.18   4.3298 0.0467155 *
## Residuals 28 169.29    6.05
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

To further optimize this model, let's look at weight and quarter mile time as functions of transmission.

```
mdlAM <- lm(mpg ~ factor(am):wt + factor(am):qsec, data = mtcars)
```

Conclusion

As can be seen in Appendix C, the multiple R-squared is 89.5%, which means that the model has 89.5% variability.

- manual transmission's effect on mpg in relation with weight, has almost twice the inverse impact compared to automatic: manual = -6.0991935 | automatic = -3.1758862. Therefore, if weight is a factor with mpg, manual transmission is better.
- manual transmision's effect on mpg in relation to quarter mile time is 1.4463757 compared to automatic transmission's 0.8337859. Therefore, if quarter mile time is a factor with mpg, manual transmission is

transmission is better. Therefore, if quarter mile time is a factor with mpg, manual transmission is better.

In conclusion, manual transmission is better for MPG.

Appendices:

Appendix A: Explore cars

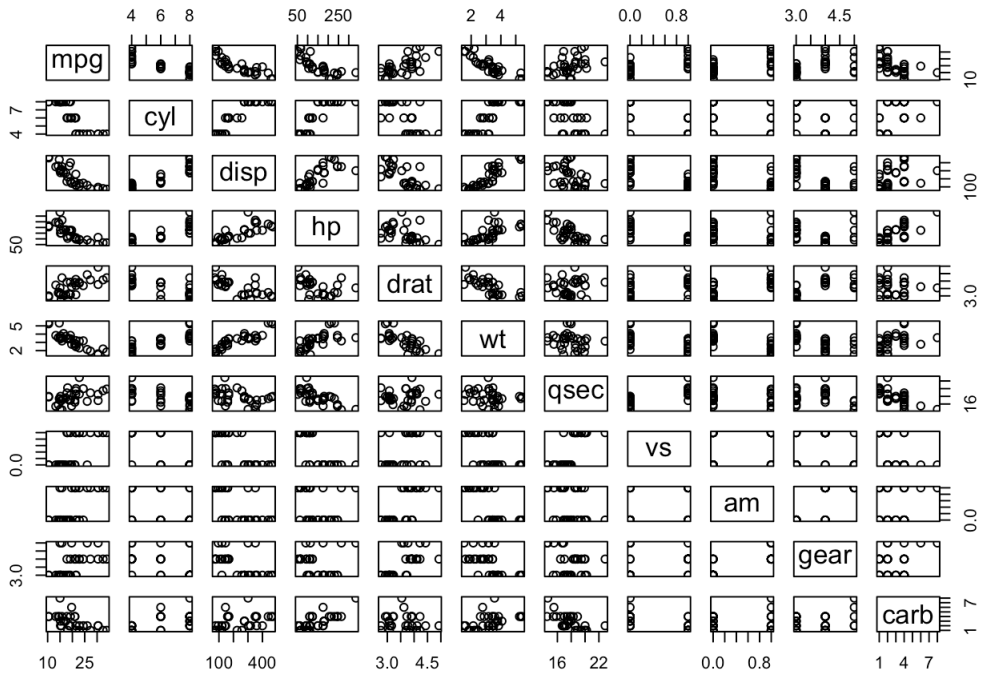
```
summary(mtcars[am == 0]) # Transmission = Automatic
```

```
##       mpg             cyl             disp             hp       
##  Min.   :10.40   Min.   :4.000   Min.   :120.1   Min.   : 62.0  
##  1st Qu.:14.95   1st Qu.:6.000   1st Qu.:196.3   1st Qu.:116.5  
##  Median :17.30   Median :8.000   Median :275.8   Median :175.0  
##  Mean   :17.15   Mean   :6.947   Mean   :290.4   Mean   :160.3  
##  3rd Qu.:19.20   3rd Qu.:8.000   3rd Qu.:360.0   3rd Qu.:192.5  
##  Max.   :24.40   Max.   :8.000   Max.   :472.0   Max.   :245.0  
##       drat             wt             qsec             vs        
##  Min.   :2.760   Min.   :2.465   Min.   :15.41   Min.   :0.0000  
##  1st Qu.:3.070   1st Qu.:3.438   1st Qu.:17.18   1st Qu.:0.0000  
##  Median :3.150   Median :3.520   Median :17.82   Median :0.0000  
##  Mean   :3.286   Mean   :3.769   Mean   :18.18   Mean   :0.3684  
##  3rd Qu.:3.695   3rd Qu.:3.842   3rd Qu.:19.17   3rd Qu.:1.0000  
##  Max.   :3.920   Max.   :5.424   Max.   :22.90   Max.   :1.0000  
##        am          gear            carb      
##  Min.   :0   Min.   :3.000   Min.   :1.000  
##  1st Qu.:0   1st Qu.:3.000   1st Qu.:2.000  
##  Median :0   Median :3.000   Median :3.000  
##  Mean   :0   Mean   :3.211   Mean   :2.737  
##  3rd Qu.:0   3rd Qu.:3.000   3rd Qu.:4.000  
##  Max.   :0   Max.   :4.000   Max.   :4.000
```

```
summary(mtcars[am == 1]) # Transmission = Manual
```

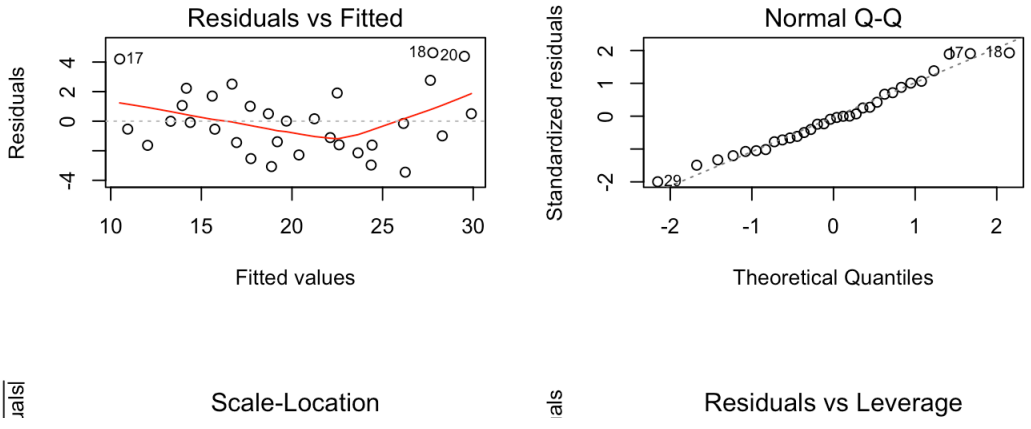
```
##       mpg             cyl             disp             hp       
##  Min.   :15.00   Min.   :4.000   Min.   : 71.1   Min.   : 52.0  
##  1st Qu.:21.00   1st Qu.:4.000   1st Qu.: 79.0   1st Qu.: 66.0  
##  Median :22.80   Median :4.000   Median :120.3   Median :109.0  
##  Mean   :24.39   Mean   :5.077   Mean   :143.5   Mean   :126.8  
##  3rd Qu.:30.40   3rd Qu.:6.000   3rd Qu.:160.0   3rd Qu.:113.0  
##  Max.   :33.90   Max.   :8.000   Max.   :351.0   Max.   :335.0  
##       drat             wt             qsec             vs        
##  Min.   :3.54   Min.   :1.513   Min.   :14.50   Min.   :0.0000  
##  1st Qu.:3.85   1st Qu.:1.935   1st Qu.:16.46   1st Qu.:0.0000  
##  Median :4.08   Median :2.320   Median :17.02   Median :1.0000  
##  Mean   :4.05   Mean   :2.411   Mean   :17.36   Mean   :0.5385  
##  3rd Qu.:4.22   3rd Qu.:2.780   3rd Qu.:18.61   3rd Qu.:1.0000  
##  Max.   :4.93   Max.   :3.570   Max.   :19.90   Max.   :1.0000  
##        am          gear            carb      
##  Min.   :1   Min.   :4.000   Min.   :1.000  
##  1st Qu.:1   1st Qu.:4.000   1st Qu.:1.000  
##  Median :1   Median :4.000   Median :2.000  
##  Mean   :1   Mean   :4.385   Mean   :2.923  
##  3rd Qu.:1   3rd Qu.:5.000   3rd Qu.:4.000  
##  Max.   :1   Max.   :5.000   Max.   :8.000
```

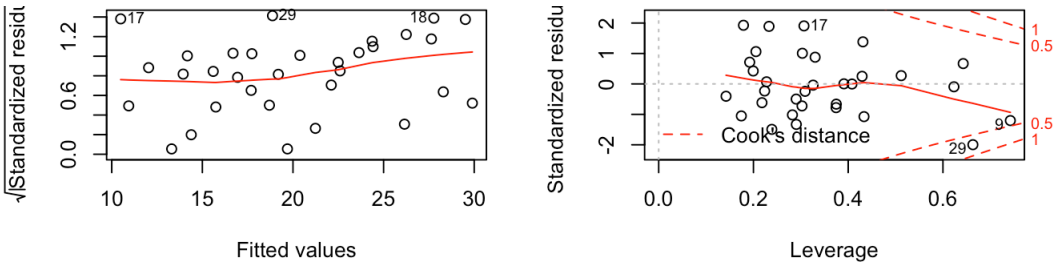
```
pairs(mpg ~ ., data = mtcars)
```



Appendix B: Model Fitting and Selection

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

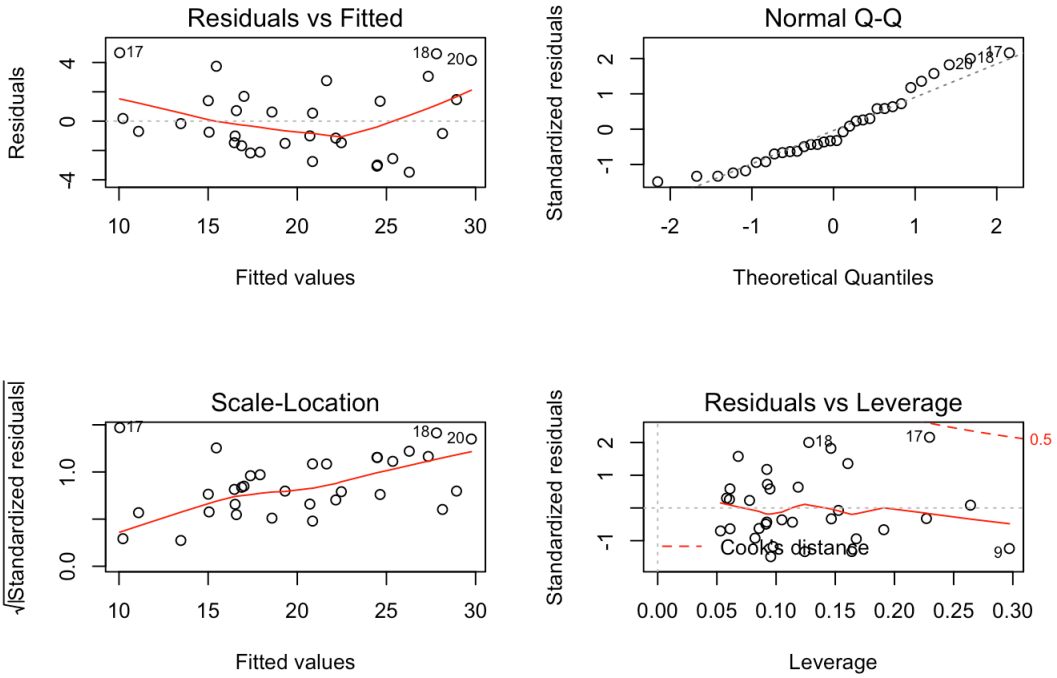





```
summary(mdlAll)
```

```
##
## Call:
## lm(formula = mpg ~ ., data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4506 -1.6044 -0.1196  1.2193  4.6271
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 12.30337   18.71788   0.657   0.5181
## cyl         -0.11144    1.04502  -0.107   0.9161
## disp         0.01334    0.01786   0.747   0.4635
## hp          -0.02148    0.02177  -0.987   0.3350
## drat         0.78711    1.63537   0.481   0.6353
## wt          -3.71530    1.89441  -1.961   0.0633 .
## qsec         0.82104    0.73084   1.123   0.2739
## vs           0.31776    2.10451   0.151   0.8814
## am           2.52023    2.05665   1.225   0.2340
## gear         0.65541    1.49326   0.439   0.6652
## carb        -0.19942    0.82875  -0.241   0.8122
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.65 on 21 degrees of freedom
## Multiple R-squared:  0.869,  Adjusted R-squared:  0.8066
## F-statistic: 13.93 on 10 and 21 DF,  p-value: 3.793e-07
```

```
plot(mdlBest)
```




```
summary(mdlBest)
```

```
##
## Call:
## lm(formula = mpg ~ wt + qsec + am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4811 -1.5555 -0.7257  1.4110  4.6610
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   9.6178     6.9596   1.382 0.177915
## wt           -3.9165     0.7112  -5.507 6.95e-06 ***
## qsec          1.2259     0.2887   4.247 0.000216 ***
## am            2.9358     1.4109   2.081 0.046716 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.459 on 28 degrees of freedom
## Multiple R-squared:  0.8497, Adjusted R-squared:  0.8336
## F-statistic: 52.75 on 3 and 28 DF,  p-value: 1.21e-11
```

Appendix C: Diagnostics

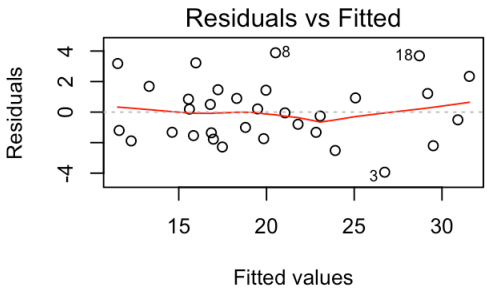
```
summary(mdlAM)
```

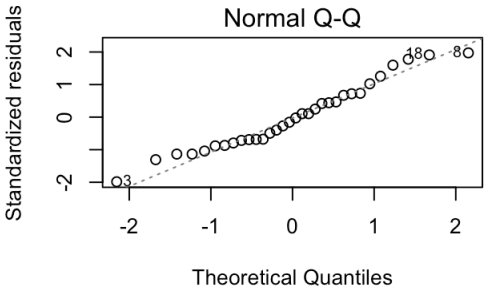
```
##
## Call:
## lm(formula = mpg ~ factor(am):wt + factor(am):qsec, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.9361 -1.4017 -0.1551  1.2695  3.8862
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)       13.9692     5.7756   2.419  0.02259 *
## factor(am)0:wt    -3.1759     0.6362  -4.992 3.11e-05 ***
## factor(am)1:wt    -6.0992     0.9685  -6.297 9.70e-07 ***
## factor(am)0:qsec   0.8338     0.2602   3.205  0.00346 **
## factor(am)1:qsec   1.4464     0.2692   5.373 1.12e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.097 on 27 degrees of freedom
## Multiple R-squared:  0.8946, Adjusted R-squared:  0.879
## F-statistic: 57.28 on 4 and 27 DF,  p-value: 8.424e-13
```

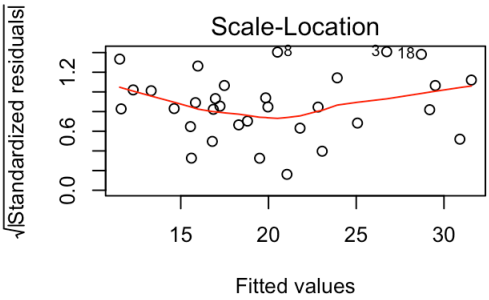
```
confint(mdlAM)
```

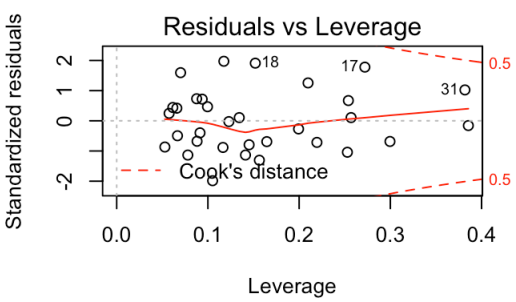
```
##                       2.5 %    97.5 %
## (Intercept)       2.1186308 25.819783
## factor(am)0:wt   -4.4813221 -1.870450
## factor(am)1:wt   -8.0864869 -4.111900
## factor(am)0:qsec  0.2999593  1.367612
## factor(am)1:qsec  0.8939972  1.998754
```

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


Residuals vs Fitted
Residuals
Fitted values
8
18
3


Normal Q-Q
Standardized residuals
Theoretical Quantiles
18
8
3


Scale-Location
√|Standardized residuals|
Fitted values
8
3
18


Residuals vs Leverage
Standardized residuals
Leverage
Cook's distance
18
17
31
0.5
0.5