Motor Trend Car Road Analysis

Glenn Kerbein

April 19, 2017

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 whether:

1. An automatic or manual transmission is better for MPG, and
2. If the difference in MPG can be quantified between automatic and manual transmissions

Data Processing

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). This is stored in the `mtcars` dataset.

```
library(datasets)
data(mtcars)
```

It consists of 32 observations on 11 variables:

1. mpg Miles per (US) gallon
2. cyl Number of cylinders
3. disp Displacement (in cubic inches)
4. hp Gross horsepower
5. drat Rear axle ratio
6. wt Weight (lb/1000)
7. qsec quarter-mile time
8. vs V/S
9. am Transmission (0 = automatic, 1 = manual)
10. gear Number of forward gears
11. carb Number of carburetors

```
summary(mtcars)
```

```
##       mpg             cyl             disp             hp
##  Min.   :10.40   Min.   :4.000   Min.   : 71.1   Min.   : 52.0
##  1st Qu.:15.43   1st Qu.:4.000   1st Qu.:120.8   1st Qu.: 96.5
##  Median :19.20   Median :6.000   Median :196.3   Median :123.0
##  Mean   :20.09   Mean   :6.188   Mean   :230.7   Mean   :146.7
##  3rd Qu.:22.80   3rd Qu.:8.000   3rd Qu.:326.0   3rd Qu.:180.0
##  Max.   :33.90   Max.   :8.000   Max.   :472.0   Max.   :335.0
##       drat             wt             qsec             vs
##  Min.   :2.760   Min.   :1.513   Min.   :14.50   Min.   :0.0000
##  1st Qu.:3.080   1st Qu.:2.581   1st Qu.:16.89   1st Qu.:0.0000
##  Median :3.695   Median :3.325   Median :17.71   Median :0.0000
##  Mean   :3.597   Mean   :3.217   Mean   :17.85   Mean   :0.4375
##  3rd Qu.:3.920   3rd Qu.:3.610   3rd Qu.:18.90   3rd Qu.:1.0000
```

```
##  Max.   :4.930   Max.   :5.424   Max.   :22.90   Max.   :1.0000
##        am              gear            carb
##  Min.   :0.0000   Min.   :3.000   Min.   :1.000
##  1st Qu.:0.0000   1st Qu.:3.000   1st Qu.:2.000
##  Median :0.0000   Median :4.000   Median :2.000
##  Mean   :0.4062   Mean   :3.688   Mean   :2.812
##  3rd Qu.:1.0000   3rd Qu.:4.000   3rd Qu.:4.000
##  Max.   :1.0000   Max.   :5.000   Max.   :8.000
```

Exploratory Data Analysis

Begin this exploration with a pairwise scatter plot between all variables (see appendix). Then, with the distribution of the dependent variable mpg, assess if it meets assumptions of regression. For this test, a histogram and kernel density would be apropos (see appendix). It appears that the data is approximately normal, or perhaps right skewed. ## Is an automatic or manual transmission better for MPG? First, some dissonance between the two transmission types. Automatic transmission MPG:

```
summary(mtcars[mtcars$am == 0,]$mpg)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   10.40   14.95   17.30   17.15   19.20   24.40
```

Manual transmission MPG:

```
summary(mtcars[mtcars$am == 1,]$mpg)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   15.00   21.00   22.80   24.39   30.40   33.90
```

See appendix for boxplot of MPG by transmission types. Ostensibly, given that the mean of manual is higher than that of automatic transmissions, we have a good hypothesis that manuals get better mileage than automatic transmissions. A t-test is appropriate here:

```
t.test(mtcars$mpg ~ mtcars$am, conf.level=0.95)
```

```
##
##  Welch Two Sample t-test
##
## data:  mtcars$mpg by mtcars$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 0 mean in group 1
##        17.14737        24.39231
```

Given that our p-value (0.001373638) is less than our tested alpha (0.05), we can reject the null hypothesis that there is no difference in MPG, and manual transmission performs better in MPG than automatic transmission, ceteris paribus.

Quantify the MPG difference between automatic and manual transmissions

In this secion we aim to quantify the MPG differences between transmission types, and attempt to account for any other variables. Create a multivariable linear regression with all variables in mtcars:

```
all <- lm(mpg ~ ., data = mtcars)
summary(all)
```

```
##
## 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
```

From the coefficients, it looks like wt is a major deciding factor. However, adding all variables may result in noise or overfitting. We must choose the best model. Using R's automatic variable selection algorithm, step, R will provide the best linear model using the AIC method.

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

```
##
## 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
```

It appears that the best fitting model includes weight, quarter-mile time, and transmission type. Weight negatively changes with MPG; quarter-mile time and transmission type positively affects MPG. Every one pound per ton weight increase will cause a decrease of roughly 4 mpg; every increase of quarter-mile time will cause an increase of 1.2 MPG; on average, manual transmission is 2.9 mpg better than automatic transmission. This model has an R squared value of approximately 0.85, meaning this model can explain 85% of variance. The model is able to explain 85% of variance. The residual plots also seems to be randomly scattered, as seen in the appendix.

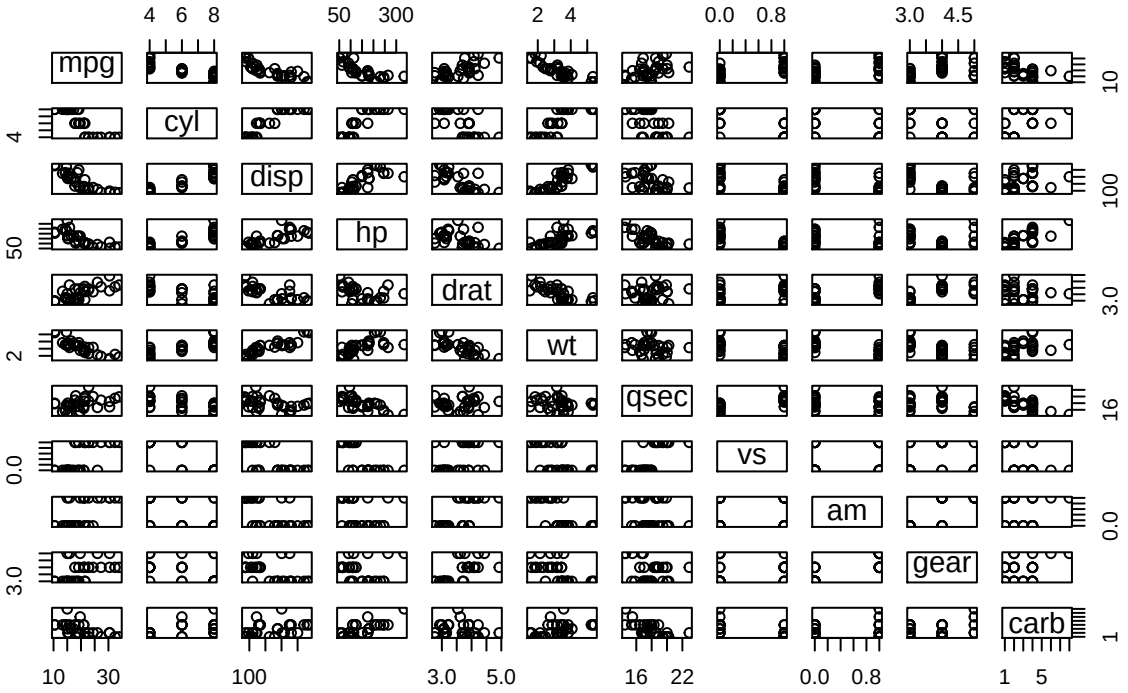
Conclusion

On average, manual transmission is better than automatic transmission by 2.9 MPG. However, transmission type is not the only factor accounting for MPG, weight, and acceleration (quarter-mile time) also needs to be considered.

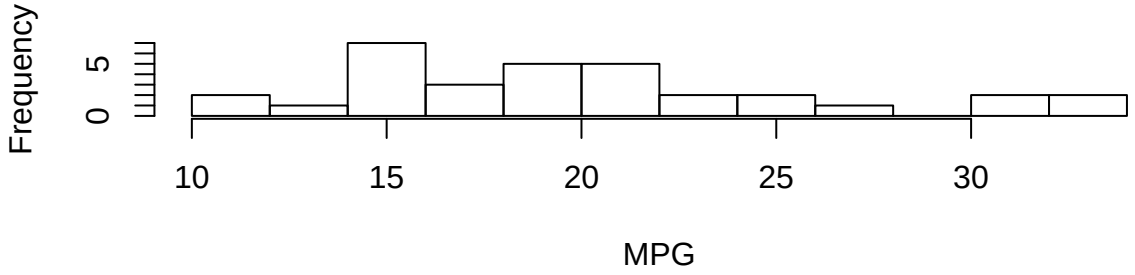
Appendix

```r
pairs(
    mtcars,
    main = "Pairwise scatter plot"
)
```

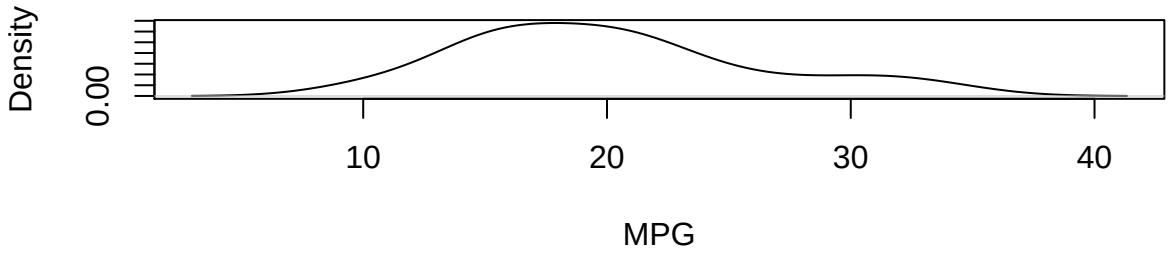
Pairwise scatter plot

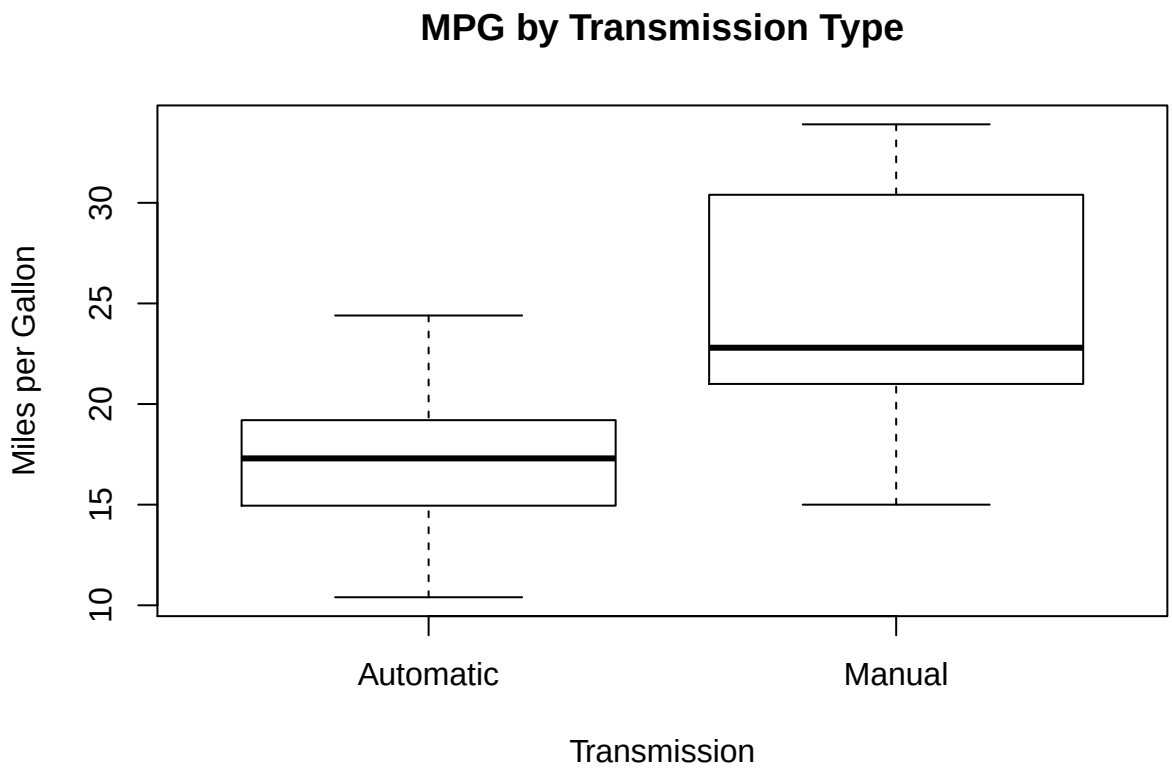


MPG histogram

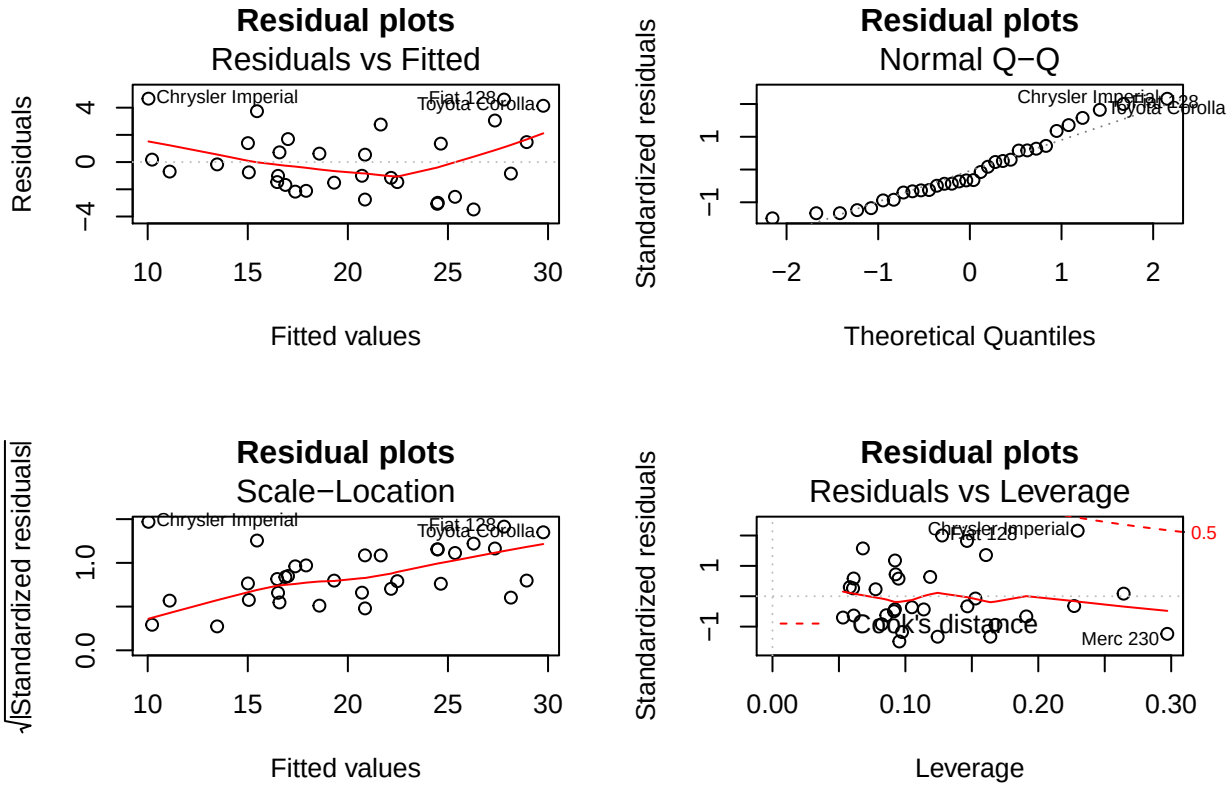


Kernel density






```
par(mfrow = c(2,2))
plot(fittedmodel, main="Residual plots")
```


Residual plots
Residuals vs Fitted
Residuals
Fitted values
Chrysler Imperial
Fiat 128
Toyota Corolla
Residual plots
Normal Q−Q
Standardized residuals
Theoretical Quantiles
Residual plots
Scale−Location
√|Standardized residuals|
Fitted values
Residual plots
Residuals vs Leverage
Standardized residuals
Leverage
Cook's distance
0.5
Merc 230