Motor Trend Cars Mileage Analysis

Dushan Yovetich

December 24, 2016

Analysis prepared for Motor Trend, an automobile industry publication. The purpose of our analysis is to explore the relationship between fuel consumption and automobile design for a set of 1973-1974 models. In particular, we are looking to:

- Determine whether an automatic or manual transmission is better for MPG
- Quantify the MPG difference between automatic and manual transmissions

Summary

There is a difference in fuel consumption between an automatic and manual transmission. Manual transmissions tended to be more fuel efficient overall, and got 2.94 miles per gallon more on average than an automatic transmission.

Motor Trend Data

A series of road tests were performed by Motor Trend in which gas mileage and other attributes were recorded for automobile models. The dataset used is an extract of road tests published in the March, April, June and July issues of the 1974 edition of the magazine.

```r
# load data
data(mtcars)
```

Our dataset includes 11 variables for a total of 32 models of automobiles. The following is a detailed description of the variables:

Variable	Description	Type
mpg	Fuel Consumption (Miles per gallon)	Continuous
cyl	Number of cylinders	Discrete
disp	Cyl displacement (i.e. Engine Size)(cubic inch)	Continuous
hp	Horsepower	Continuous
drat	Final Drive Ratio	Continuous
wt	Weight (thousand pounds)	Continuous
qsec	Quarter mile time (Seconds)	Continuous
vs	Engine Shape	Discrete (V = 0, Straight = 1)
am	Transmission Type	Discrete (Auto = 0, Manual = 1)
gear	Number of Transmission Speeds	Discrete
carb	Number of Carboretur Barrels	Discrete

R Packages

The following are additional R packages used in preparation of our analysis.

```r
library(ggplot2)
library(dplyr)
library(GGally)
```

```r
library(gridExtra)
```

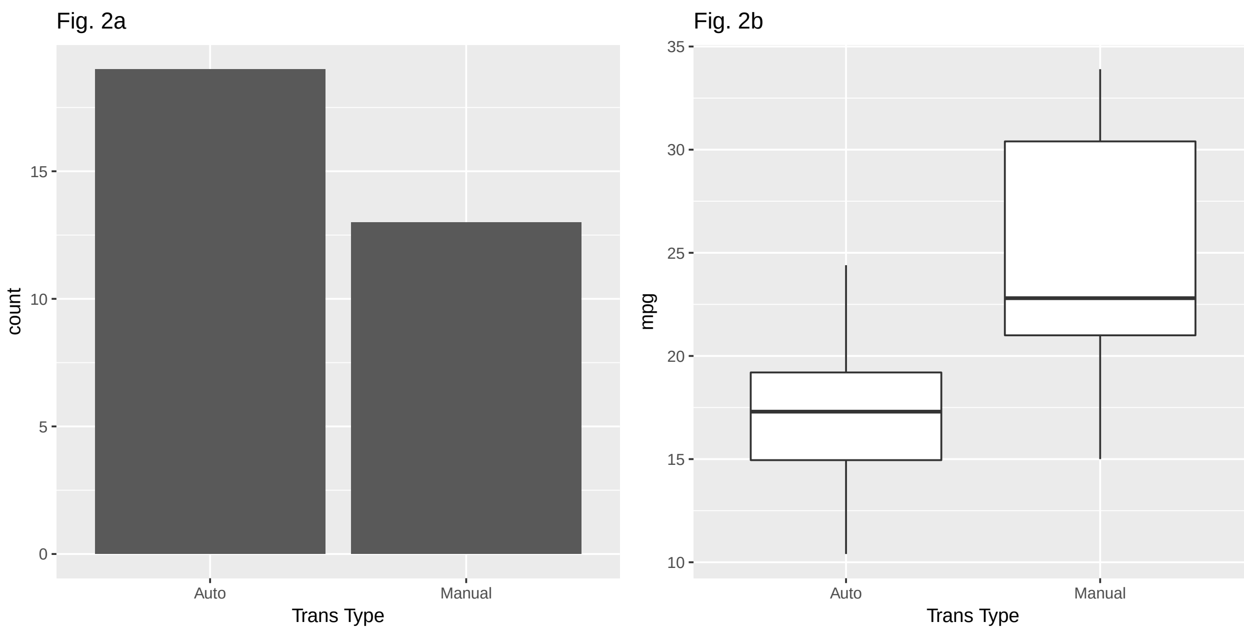
Fuel Consumption By Transmission Type

We look at the population for each transmission. **Fig. 2a** is a count of automobile models by transmission. There is a higher incidence of models with automatic versus manual transmissions. We next look at the relationship between transmissions and fuel consumption. **Fig. 2b** is a plot of mpg by transmission. Our figure shows manual transmissions on average have higher miles per gallon. However, further testing is needed to confirm there is a difference.

```r
plot1 <- qplot(x = factor(am,labels = c("Auto","Manual")),data = mtcars, geom="bar",
               main = "Fig. 2a", xlab = "Trans Type")

plot2 <- qplot(x = factor(am,labels = c("Auto","Manual")), y = mpg, data = mtcars,
               geom = "boxplot", main = "Fig. 2b", xlab = "Trans Type")

grid.arrange(plot1,plot2,nrow = 1, ncol = 2)
```



Test of Two Transmission Types

To determine whether the difference in fuel consumption for transmission type is significant, we subset our dataset into two sets based on transmission and perform a Student's t-Test of:

- H0 = 0 no difference between transmission types
- Ha <> 0 a difference exists between transmission types

We assume a 95% confidence level ($|p| > 0.05$), and the variances for the two sets are not equal.

```r
mt <- with(mtcars,t.test(mpg[am==0],mpg[am==1]))
print(mt)
```

```
##
##  Welch Two Sample t-test
```

```
##
## data:  mpg[am == 0] and mpg[am == 1]
## 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 of x mean of y
##  17.14737  24.39231
```

Our p-value, 0.0013736, is significant with a 95% confidence level. We therefore fail to reject the alternate hypothesis (Ha). In conclusion, a difference exists between automatic and manual transmissions with regard to fuel consumption.

Quantifying Fuel Consumption Differences Between Transmission Types

Our analysis next attempts to quantify differences in fuel consumption by transmission. We begin with a simple linear regression model using fuel consumption, `mpg`, as the dependent variable and transmission, `am`, as our independent variable. A 95% confidence level ($|p| < 0.05$) is assumed.

```r
fit <- lm(mpg ~ factor(am), mtcars)
summary(fit)$coeff
```

```
##              Estimate Std. Error   t value                Pr(>|t|)
## (Intercept) 17.147368   1.124603 15.247492 0.000000000000001133983
## factor(am)1  7.244939   1.764422  4.106127 0.000285020743935067769
```

The independent variable, `am`, is significant with a p-value of 0.000285. However, the percentage of total variability explained by our model is low at 0.3597989. Confounding variables may need to be considered.

Multivariate Linear Regression

We attempt a multivariate model with all variables in our dataset. Again, assuming a 95% confidence level ($|p| > 0.05$).

```r
fit2 <- lm(mpg~., mtcars)
summary(fit2)$coeff
```

```
##                Estimate  Std. Error    t value   Pr(>|t|)
## (Intercept) 12.30337416 18.71788443  0.6573058 0.51812440
## cyl         -0.11144048  1.04502336 -0.1066392 0.91608738
## disp         0.01333524  0.01785750  0.7467585 0.46348865
## hp          -0.02148212  0.02176858 -0.9868407 0.33495531
## drat         0.78711097  1.63537307  0.4813036 0.63527790
## wt          -3.71530393  1.89441430 -1.9611887 0.06325215
## qsec         0.82104075  0.73084480  1.1234133 0.27394127
## vs           0.31776281  2.10450861  0.1509915 0.88142347
## am           2.52022689  2.05665055  1.2254035 0.23398971
## gear         0.65541302  1.49325996  0.4389142 0.66520643
## carb        -0.19941925  0.82875250 -0.2406258 0.81217871
```

The percentage of total variability now explained is 0.8690158, an improvement over the simple model. We observe the variable `wt` is shown to be significant, as well as possibly `qsec` and `hp`. However, `qsec` and `hp` have a correlation of -0.7082234. We therefore choose to omit `hp` from our model, since `qsec` was shown to be more significant. The result is a model with `mpg` as the dependent outcome, and the independent predictors `am`, `wt`, and `qsec`. We apply our model, assuming a 95% confidence level ($|p| < 0.05$).

```
fit3 <- lm(mpg~am+wt+qsec, data = mtcars)
summary(fit3)$coeff
```

```
##              Estimate Std. Error   t value       Pr(>|t|)
## (Intercept)  9.617781  6.9595930  1.381946 0.177915165459
## am           2.935837  1.4109045  2.080819 0.046715509919
## wt          -3.916504  0.7112016 -5.506882 0.000006952711
## qsec         1.225886  0.2886696  4.246676 0.000216173705
```

The percentage of total variability now explained by the linear relationship is 0.8496636. Further, the p-values of the independent predictors are all significant. We confirm this is indeed the most appropriate model and review of our diagnostic plots show no significant systematic patterns in our residuals, or highly influential or leveraged observations, that would cause us to dismiss our selected model. (See *Appendix, Stepwise Regression* and *Appendix, Model Diagnostics*). In conclusion, there is a **2.9358372** mpg difference between automatic and manual transmissions.

Appendix

Stepwise Regression

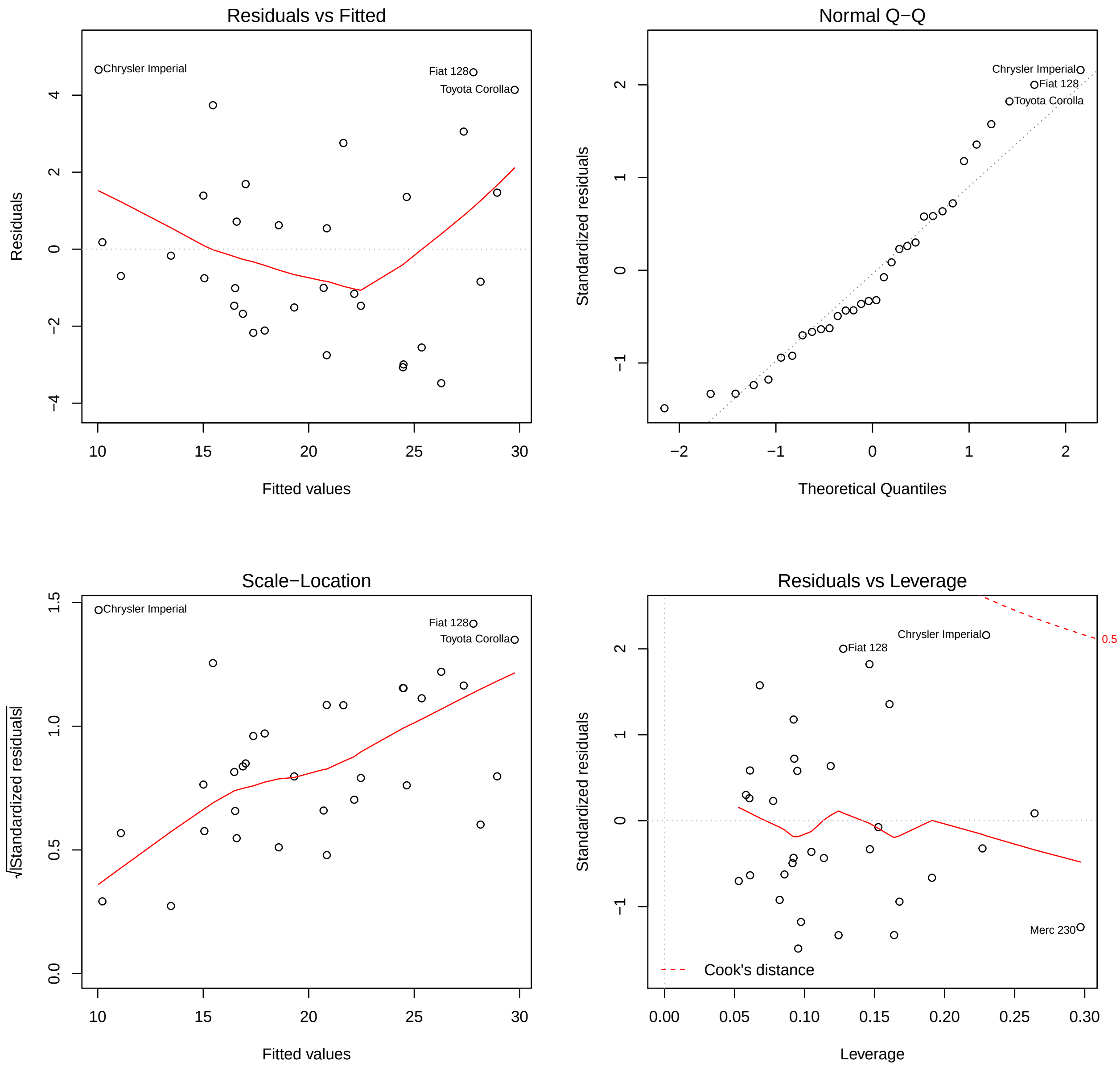
```
step.fit <- step(lm(mpg~.,data = mtcars), trace = 0)
summary(step.fit)
```

```
##
## 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 0.00000695 ***
## 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: 0.0000000000121
```

Model Diagnostics

```
# plot of diagnostics
par(mfrow=c(2,2))
```

```r
plot(fit3)
```



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