Programming Assigment 1

The Effect of Transmission Type on MPG

Author: Zdravka Cankova

Data: mtcars dataset built into R

Source: Henderson and Velleman (1981), Building multiple regression models interactively. Biometrics, 37, 391–411.

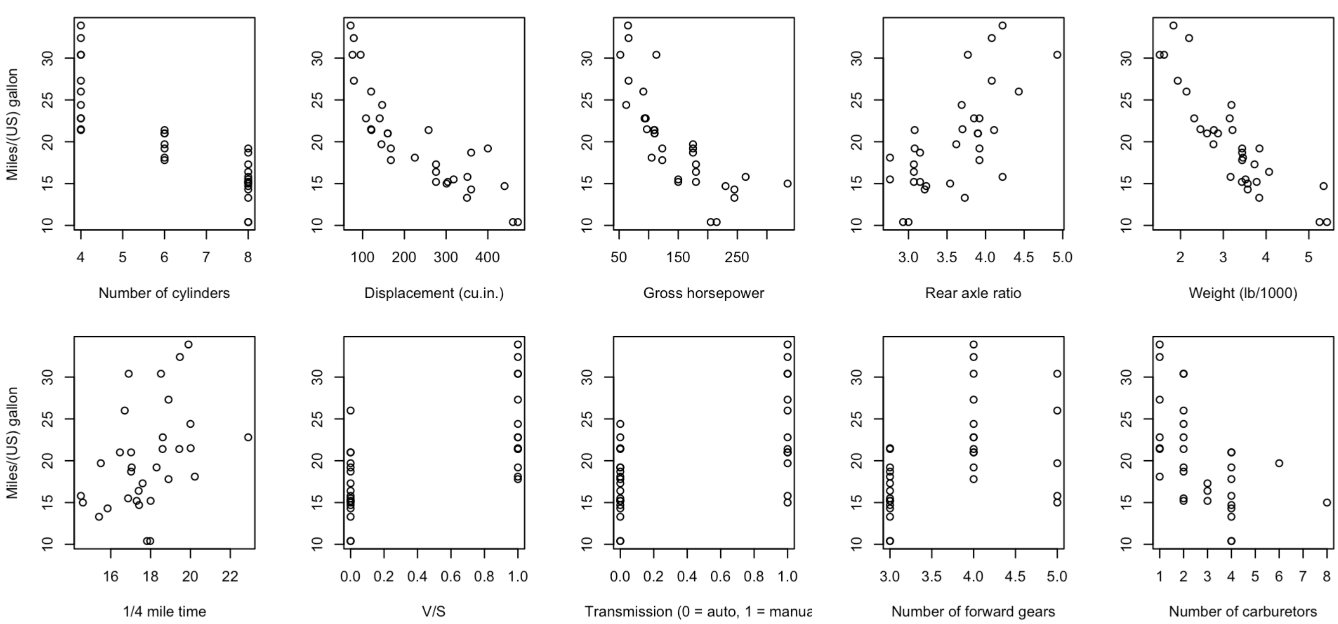
Overview

The purpose of this project was to use the “mtcars” dataset provided with R to quantify the MPG difference between manual and automatic transmission and determine which type is better for MPG. Multivriate regression analysis was used to show that manual transmission is better for MPG, resulting in a 4.16 mpg increase when compared to automatic.

Data Analysis

First, I looked at the scatterplots of MPG vs. the remaining variables to check if there are any obvious relationships (code shown in Appendix).

Figure 1: Scatterplots Exploring the Changes in MPG Based on Other Variables



After examining the plots, it appears that most variables seem to be correlated to miles/gallon, except maybe the rear axle ratio, 1/4 mile time and number of forward gears.

As evidenced in the scatterplots, several variables have clear levels (number of cylinders, V/S, transmission, number of forward gears, and number of carburetors), and so these were converted to factor variables (code not shown).

Then I used nested model testing to determine which variables other than transmission have a significant ($p < 0.05$) contribution to the model (code shown in Appendix).

```
## Analysis of Variance Table
##
## Model  1: mpg ~ am
## Model  2: mpg ~ am + cyl
## Model  3: mpg ~ am + cyl + disp
## Model  4: mpg ~ am + cyl + disp + hp
## Model  5: mpg ~ am + cyl + disp + hp + drat
## Model  6: mpg ~ am + cyl + disp + hp + drat + wt
## Model  7: mpg ~ am + cyl + disp + hp + drat + wt + qsec
## Model  8: mpg ~ am + cyl + disp + hp + drat + wt + qsec + vs
## Model  9: mpg ~ am + cyl + disp + hp + drat + wt + qsec + vs + gear
## Model 10: mpg ~ am + cyl + disp + hp + drat + wt + qsec + vs + gear + carb
##    Res.Df    RSS Df Sum of Sq       F   Pr(>F)
## 1      30 720.90
## 2      28 264.50  2    456.40 28.4297 7.89e-06 ***
## 3      27 230.46  1     34.04  4.2402  0.05728 .
## 4      26 183.04  1     47.42  5.9078  0.02809 *
## 5      25 182.38  1      0.66  0.0820  0.77855
## 6      24 150.10  1     32.28  4.0216  0.06331 .
## 7      23 141.21  1      8.89  1.1081  0.30916
## 8      22 139.02  1      2.18  0.2719  0.60964
## 9      20 134.00  2      5.02  0.3128  0.73606
## 10     15 120.40  5     13.60  0.3388  0.88144
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Besides transmission type, only the number of cylinders and gross horsepower contributed significantly ($p < 0.05$) to the regression model. After inlcuding only these variables, I obtained the following model:

```
fitFinal <- lm(mpg ~ am + cyl + hp, data = mtcars)
summary(fitFinal)$coefficients
```

```
##                Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 27.29589929 1.42394266 19.169240 2.977929e-17
## am1          4.15785647 1.25655004  3.308946 2.659806e-03
## cyl6        -3.92457850 1.53751488 -2.552547 1.666330e-02
## cyl8        -3.53341392 2.50278816 -1.411791 1.694330e-01
## hp          -0.04424394 0.01457563 -3.035473 5.266203e-03
```

Based on the beta coefficient summary, I can conlcude that **manual transmission is significantly better for MPG** ($p = 0.003$), leading to an average **4.16 mpg increase** when compared to automatic transmission.

The Appendix contains several plots that demonstate the validity of the regression model. The residuals vs. fitted and scale-location plots show that the residuals are identically and independently distributed (iid). The Q-Q plot shows that the residuals are not skewed and there are no outliers. The residuals vs. leverage plot also demonstrates that there are no outliers or highly influential points. These results indicate that this model is an accurate representation of the true relationship between the data.

Appendix

The code used to generate the exploratory scatterplots shown in Figure 1 is below:

```
data(mtcars)
par(mfrow = c(2, 5), mar = c(4, 4, 2, 1), oma = c(0, 0, 2, 0))
with(mtcars, {
      plot(cyl, mpg,
           xlab = "Number of cylinders",
           ylab = "Miles/(US) gallon")
      plot(disp, mpg,
           xlab = "Displacement (cu.in.)",
           ylab = "")
      plot(hp, mpg,
           xlab = "Gross horsepower",
           ylab = "")
      plot(drat, mpg,
           xlab = "Rear axle ratio",
           ylab = "")
      plot(wt, mpg,
           xlab = "Weight (lb/1000)",
           ylab = "")
      plot(qsec, mpg,
           xlab = "1/4 mile time",
           ylab = "Miles/(US) gallon")
      plot(vs, mpg,
           xlab = "V/S",
           ylab = "")
      plot(am, mpg,
           xlab = "Transmission (0 = auto, 1 = manual)",
           ylab = "")
      plot(gear, mpg,
           xlab = "Number of forward gears",
           ylab = "")
      plot(carb, mpg,
           xlab = "Number of carburetors",
           ylab = "")})
mtext("Figure 1: Scatterplots Exploring the Changes in MPG Based on Other Variables", font =
2, adj = 0, outer = TRUE)
```

The code used for nested model testing is shown below:

```
fitAM <- lm(mpg ~ am, data = mtcars)
fitCyl <- update(fitAM, mpg ~ am + cyl)
fitDisp <- update(fitAM, mpg ~ am + cyl + disp)
fitHp <- update(fitAM, mpg ~ am + cyl + disp + hp)
fitDrat <- update(fitAM, mpg ~ am + cyl + disp + hp + drat)
fitWt <- update(fitAM, mpg ~ am + cyl + disp + hp + drat + wt)
fitQsec <- update(fitAM, mpg ~ am + cyl + disp + hp + drat + wt + qsec)
fitVs <- update(fitAM, mpg ~ am + cyl + disp + hp + drat + wt + qsec + vs)
fitGear <- update(fitAM, mpg ~ am + cyl + disp + hp + drat + wt + qsec + vs + gear)
fitCarb <- update(fitAM, mpg ~ am + cyl + disp + hp + drat + wt + qsec + vs + gear + carb)
anova(fitAM, fitCyl, fitDisp, fitHp, fitDrat, fitWt, fitQsec, fitVs, fitGear, fitCarb)
```

The residuals and other diagnostic plots for the final regression model are shown in Figure 3 below:

```
par(mfrow = c(2, 2), mar = c(4, 4, 2, 1), oma = c(0, 0, 2, 0))
plot(fitFinal, sub.caption = NA)
mtext("Figure 2: Residual Plots and Diagnostics", font = 2, adj = 0, outer = TRUE)
```

Figure 2: Residual Plots and Diagnostics