

Prediction Assignment Writeup

pvv88

Executive Summary

The data from mtcars will be analyzed to determine the relationship between the miles per gallon (mpg) and the transmission of the car. We use regression models and exploratory data analyses to identify how automatic (am = 0) and manual (am = 1) transmissions features affect the MPG feature. The t-test shows that there is a performance difference between cars with automatic and manual transmission.

Data Processing

Load the data from mtcars and changed some variables (cyl, vs, gear, carb, and am) from numeric to factor.

```
data(mtcars)
mtcars$am = factor(mtcars$am, label = c("Automatic", "Manual"))
mtcars$cyl = as.factor(mtcars$cyl)
mtcars$gear = as.factor(mtcars$gear)
mtcars$carb = as.factor(mtcars$carb)
mtcars$vs = as.factor(mtcars$vs)
auto = subset(mtcars, mtcars$am == "Automatic")
manu = subset(mtcars, mtcars$am == "Manual")
```

Exploratory Analysis

Plots were created to show how the data is distributed amongst transmission and mpg (Appendix, "MPG vs. Transmission") and scatterplot matrix (Appendix, "Scatterplot Matrix"). There is a clear distinction between automatic transmission and manual transmission vehicle's MPG. Manual cars have a higher maximum mpg and minimum mpg compared to automatic cars.

Hypothesis Testing

```
t.test(manu$mpg, auto$mpg)

##
##  Welch Two Sample t-test
##
## data:  manu$mpg and auto$mpg
## 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:
##   3.209684 11.280194
## sample estimates:
```

```
## mean of x mean of y
## 24.39231 17.14737
```

T test shows low p-value at 0.003174 which indicates difference between manual and automatic transmissions.

Regression Analysis

Linear Regression

Where am is the predictor and mpg is the output.

```
fmodel<-lm(mpg~ am, data=mtcars)
summary(fmodel)
```

It is observed that the p-value and adjusted R-squared is less than 0.05 which is not of significance.

Multivariate Regression

```
fmodel<-lm(mpg~., data=mtcars)
bmodel<-step(fmodel, direction='both')
summary(bmodel)
```

bmodel contains the type of transmission in it, where the manual transmission has almost thrice the mileage per gallon of fuel than an automatic transmission car, while all the other variables are constant. This model explains an adjusted value of 0.8401 or 84.01% of variance. The coefficient of am1 (manual transmission) is about 1.8 which means that it has a greater weight in determining the mpg. The residual plots are shown in (Appendix, "Residual Plots of the bmodel").

Conclusion

Basing the adjusted-R values between linear (33.8%) and multivariate regression(84%), multivariate regression is the better model.

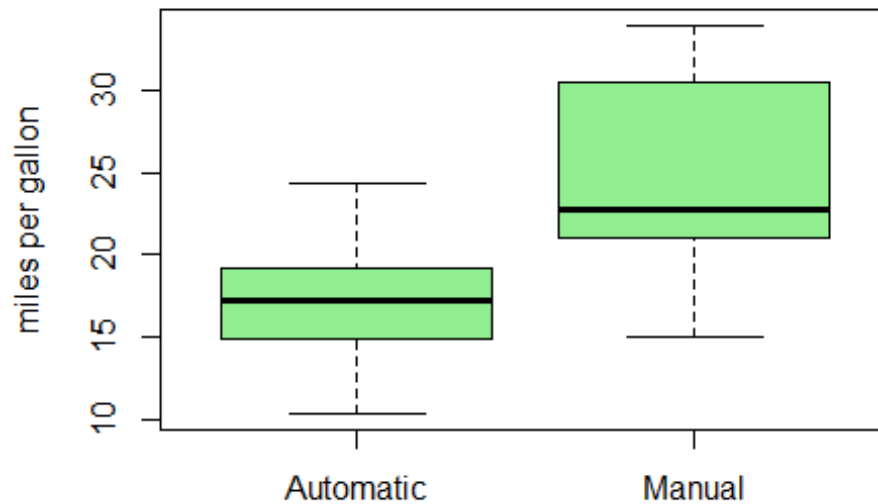
In addition, A 95% T test of the transmission of the cars to mpg shows that their means are significantly different. We therefore conclude that Manual transmission have better MPG compared to Automatic transmission.

Appendix

1. MPG vs. Transmission

```
boxplot(data = mtcars,
        mpg ~ am,
        xlab = "(am = 0) and manual (am = 1) Transmissions",
        ylab = "miles per gallon",
        main = "Boxplot of MPG vs. Transmission",
        col = "lightgreen")
```

Boxplot of MPG vs. Transmission



(am = 0) and manual (am = 1) Transmissions

2. Regression Analysis

```
## Start: AIC=76.24
## mpg ~ cyl + disp + hp + drat + wt + qsec + vs + am + gear + carb
##
##      Df Sum of Sq  RSS   AIC
## - carb  5   19.0251 146.56 70.693
## - gear  2    2.2878 129.82 72.813
## - qsec  1    0.6755 128.21 74.413
## - am    1    0.7252 128.25 74.425
## - vs    1    1.1989 128.73 74.543
## - cyl    1    3.8047 131.33 75.184
## - drat   1    4.6614 132.19 75.392
## <none>                127.53 76.244
## - disp   1   16.3306 143.86 78.099
## - hp     1   18.6718 146.20 78.616
## - wt     1   26.8119 154.34 80.350
##
## Step: AIC=70.69
## mpg ~ cyl + disp + hp + drat + wt + qsec + vs + am + gear
##
##      Df Sum of Sq  RSS   AIC
## - gear  2    2.318 148.87 67.195
## - cyl   1    0.011 146.57 68.696
## - vs    1    0.676 147.23 68.840
## - drat  1    2.654 149.21 69.267
## - disp  1    6.233 152.79 70.026
```

```

## <none>          146.56 70.693
## - qsec  1      11.682 158.24 71.147
## - am    1      12.102 158.66 71.232
## - hp    1      13.334 159.89 71.480
## - wt    1      37.870 184.42 76.048
## + carb  5      19.025 127.53 76.244
##
## Step:  AIC=67.2
## mpg ~ cyl + disp + hp + drat + wt + qsec + vs + am
##
##          Df Sum of Sq  RSS   AIC
## - vs      1      0.217 149.09 65.242
## - cyl      1      0.575 149.45 65.319
## - drat     1      1.886 150.76 65.598
## <none>          148.87 67.195
## - disp     1      9.678 158.55 67.211
## - qsec     1      9.782 158.65 67.232
## - hp       1     11.296 160.17 67.536
## - am       1     14.474 163.35 68.165
## + gear     2      2.318 146.56 70.693
## + carb     5     19.056 129.82 72.813
## - wt       1     60.737 209.61 76.144
##
## Step:  AIC=65.24
## mpg ~ cyl + disp + hp + drat + wt + qsec + am
##
##          Df Sum of Sq  RSS   AIC
## - cyl      1      1.003 150.09 63.457
## - drat     1      1.901 150.99 63.648
## - disp     1      9.548 158.64 65.228
## <none>          149.09 65.242
## - hp       1     11.366 160.46 65.593
## - qsec     1     13.344 162.43 65.985
## - am       1     14.387 163.48 66.190
## + vs       1      0.217 148.87 67.195
## + gear     2      1.859 147.23 68.840
## + carb     5     16.534 132.56 71.481
## - wt       1     65.175 214.26 74.847
##
## Step:  AIC=63.46
## mpg ~ disp + hp + drat + wt + qsec + am
##
##          Df Sum of Sq  RSS   AIC
## - drat     1      3.345 153.44 62.162
## - disp     1      8.545 158.64 63.229
## <none>          150.09 63.457
## - hp       1     13.285 163.38 64.171
## + cyl      1      1.003 149.09 65.242
## + vs       1      0.645 149.45 65.319
## - am       1     20.036 170.13 65.466

```

```

## - qsec 1 25.574 175.67 66.491
## + gear 2 2.779 147.31 66.859
## + carb 5 11.674 138.42 70.866
## - wt 1 67.572 217.66 73.351
##
## Step: AIC=62.16
## mpg ~ disp + hp + wt + qsec + am
##
## Df Sum of Sq RSS AIC
## - disp 1 6.629 160.07 61.515
## <none> 153.44 62.162
## - hp 1 12.572 166.01 62.682
## + drat 1 3.345 150.09 63.457
## + cyl 1 2.447 150.99 63.648
## + vs 1 1.121 152.32 63.927
## - qsec 1 26.470 179.91 65.255
## + gear 2 3.038 150.40 65.522
## - am 1 32.198 185.63 66.258
## + carb 5 3.965 149.47 71.324
## - wt 1 69.043 222.48 72.051
##
## Step: AIC=61.52
## mpg ~ hp + wt + qsec + am
##
## Df Sum of Sq RSS AIC
## - hp 1 9.219 169.29 61.307
## <none> 160.07 61.515
## + disp 1 6.629 153.44 62.162
## + drat 1 1.428 158.64 63.229
## - qsec 1 20.225 180.29 63.323
## + cyl 1 0.249 159.82 63.465
## + vs 1 0.249 159.82 63.466
## - am 1 25.993 186.06 64.331
## + gear 2 1.764 158.30 65.161
## + carb 5 6.393 153.67 70.211
## - wt 1 78.494 238.56 72.284
##
## Step: AIC=61.31
## mpg ~ wt + qsec + am
##
## Df Sum of Sq RSS AIC
## <none> 169.29 61.307
## + hp 1 9.219 160.07 61.515
## + disp 1 3.276 166.01 62.682
## + cyl 1 1.501 167.78 63.022
## + drat 1 1.400 167.89 63.042
## + vs 1 0.000 169.29 63.307
## - am 1 26.178 195.46 63.908
## + gear 2 0.185 169.10 65.272
## + carb 5 10.999 158.29 69.158

```

```

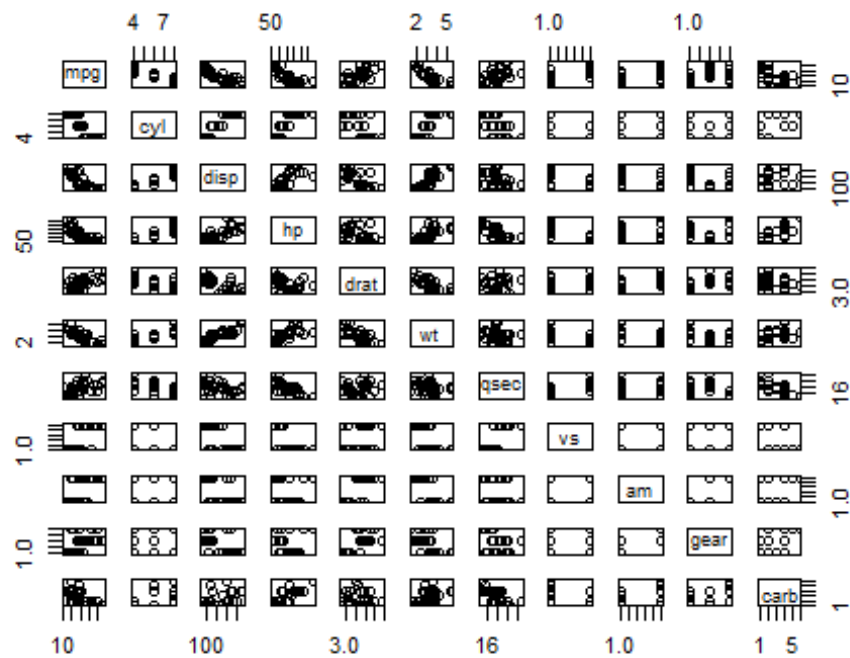
## - qsec 1 109.034 278.32 75.217
## - wt 1 183.347 352.63 82.790

##
## 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 ***
## amManual     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

##              2.5 %    97.5 %
## (Intercept) -4.63829946 23.873860
## wt          -5.37333423 -2.459673
## qsec         0.63457320  1.817199
## amManual     0.04573031  5.825944

```

3. Scatterplot Matrix



4. Residual Plots of the bmodel

```
## Loading required package: ggplot2
```

