

Untitled

Abhay Singh

24/02/2021

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”

```
data(mtcars)
```

```
#See appendix for exploratory analysis.
```

```
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
```

From the summary results I can correctly parse the data to create the regression model.

```
mtcars$drat <- factor(mtcars$drat)
mtcars$cyl <- factor(mtcars$cyl)
mtcars$vs <- factor(mtcars$vs)
mtcars$gear <- factor(mtcars$gear)
mtcars$carb <- factor(mtcars$carb)
mtcars$am <- factor(mtcars$am, labels=c('Automatic', 'Manual'))
```

REGRESSION MODEL

Compare the full model to

#Result shown in the Appendix

```
full.model <- lm(mpg ~ ., data = mtcars)
best.model <- step(full.model, direction = "backward")
```

```
## Start: AIC=25.34
## mpg ~ cyl + disp + hp + drat + wt + qsec + vs + am + gear + carb
##
##
## Step: AIC=25.34
## mpg ~ cyl + disp + hp + drat + wt + qsec + vs + am + gear
##
##
## Step: AIC=25.34
## mpg ~ cyl + disp + hp + drat + wt + qsec + vs + gear
##
##
## Step: AIC=25.34
## mpg ~ cyl + disp + hp + drat + wt + qsec + gear
##
##      Df Sum of Sq    RSS    AIC
## - qsec  1      0.083 11.615 23.571
## <none>                 11.532 25.340
## - wt    1      2.963 14.495 30.658
## - disp  1      4.773 16.305 34.423
## - hp    1      5.044 16.576 34.951
## - gear  1      5.984 17.516 36.715
## - cyl   2      9.582 21.113 40.694
## - drat 20     135.570 147.102 66.812
##
## Step: AIC=23.57
## mpg ~ cyl + disp + hp + drat + wt + gear
##
##      Df Sum of Sq    RSS    AIC
## <none>                 11.615 23.571
## - wt    1      3.569 15.184 30.144
## - disp  1      4.873 16.489 32.782
## - hp    1      5.169 16.784 33.350
## - gear  1      6.222 17.838 35.299
## - cyl   2     12.344 23.959 42.740
## - drat 20     139.637 151.253 65.703
```

#Result shown in the Appendix

```
summary(best.model)
```

```
##
## Call:
## lm(formula = mpg ~ cyl + disp + hp + drat + wt + gear, data = mtcars)
```

```

##
## Residuals:
##      Mazda RX4      Mazda RX4 Wag      Datsun 710      Hornet 4 Drive
##      6.305e-01      -6.305e-01      2.776e-17      -1.205e-01
##      Hornet Sportabout      Valiant      Duster 360      Merc 240D
##      -9.637e-02      1.205e-01      1.665e-16      -3.053e-16
##      Merc 230      Merc 280      Merc 280C      Merc 450SE
##      1.305e-15      7.000e-01      -7.000e-01      -9.385e-01
##      Merc 450SL      Merc 450SLC      Cadillac Fleetwood      Lincoln Continental
##      1.643e+00      -7.044e-01      2.776e-17      -5.551e-17
##      Chrysler Imperial      Fiat 128      Honda Civic      Toyota Corolla
##      0.000e+00      1.686e+00      -3.608e-16      -4.718e-16
##      Toyota Corona      Dodge Challenger      AMC Javelin      Camaro Z28
##      -1.943e-16      -1.205e-01      9.637e-02      -2.776e-16
##      Pontiac Firebird      Fiat X1-9      Porsche 914-2      Lotus Europa
##      1.205e-01      -1.686e+00      1.388e-16      0.000e+00
##      Ford Pantera L      Ferrari Dino      Maserati Bora      Volvo 142E
##      8.327e-17      0.000e+00      5.551e-17      -2.498e-16
##
## Coefficients: (1 not defined because of singularities)
##      Estimate Std. Error t value Pr(>|t|)
## (Intercept)  31.038    20.528   1.512  0.2051
## cyl6         -59.745    40.381  -1.480  0.2131
## cyl8        -79.839    49.838  -1.602  0.1844
## disp        -1.389     1.072  -1.295  0.2649
## hp           3.259     2.443   1.334  0.2530
## drat2.93     20.944    27.242   0.769  0.4849
## drat3       -29.174    13.566  -2.151  0.0979
## drat3.07    -157.389   119.891  -1.313  0.2595
## drat3.08     34.305    25.051   1.369  0.2427
## drat3.15    -19.546    16.086  -1.215  0.2911
## drat3.21   -252.792   187.756  -1.346  0.2494
## drat3.23   -101.149    69.620  -1.453  0.2199
## drat3.54   -627.339   470.801  -1.332  0.2535
## drat3.62   -334.116   254.614  -1.312  0.2597
## drat3.69   -225.129   157.636  -1.428  0.2264
## drat3.7    -170.968   130.460  -1.311  0.2602
## drat3.73   -269.020   199.065  -1.351  0.2479
## drat3.77   -244.231   194.278  -1.257  0.2771
## drat3.85   -377.210   272.852  -1.382  0.2390
## drat3.9    -304.535   219.304  -1.389  0.2373
## drat3.92   -342.264   244.439  -1.400  0.2340
## drat4.08   -321.424   237.585  -1.353  0.2475
## drat4.11   -414.963   299.015  -1.388  0.2375
## drat4.22   -323.732   242.931  -1.333  0.2535
## drat4.43   -145.031   114.904  -1.262  0.2755
## drat4.93   -277.391   205.793  -1.348  0.2490
## wt           4.945     4.461   1.109  0.3298
## gear4       204.483   139.689   1.464  0.2171
## gear5        NA         NA      NA      NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.704 on 4 degrees of freedom

```

```
## Multiple R-squared:  0.9897, Adjusted R-squared:  0.9201
## F-statistic: 14.21 on 27 and 4 DF,  p-value: 0.009558
```

This procedure determines that the best model includes the cyl6, cyl8, hp, wt, and amManual variables (overall p-value<0.001). The adjusted R-squared indicates that about 84% of the variance is explained by the final model. Moreover, the output of this model suggests that mpg decreases with respect to cylinders (-3.03 and -2.16 for cyl6 and cyl8, respectively), horsepower (-0.03), and weight (for every 1,000lb, by -2.5). On the other hand, mpg increases with respect to having a manual transmission (by 1.8). Residual plots (see appendix) suggest that some transformation may be necessary to achieve linearity.

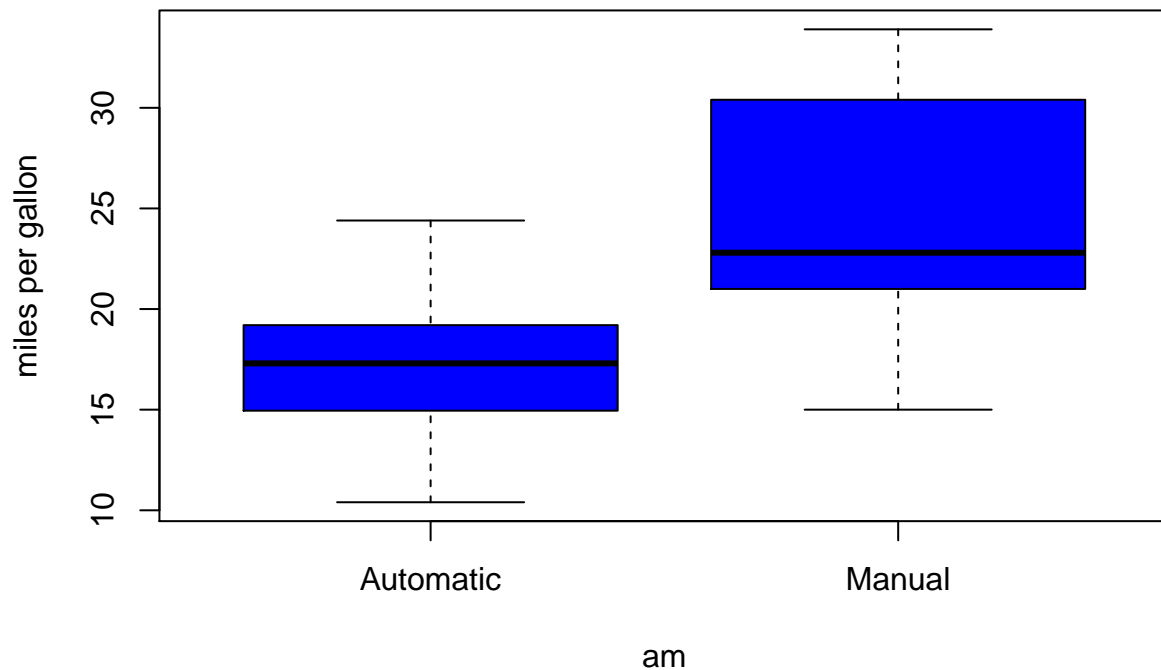
A test is then run to determine the difference in mpg between the automatic and manual transmissions.

```
t.test(mpg ~ am, data = mtcars)
```

```
##
## Welch Two Sample t-test
##
## data: mpg by 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 Automatic    mean in group Manual
##           17.14737           24.39231
```

Result shown in the Appendix

```
boxplot(mpg ~ am, data = mtcars, col = "blue", ylab = "miles per gallon")
```



The boxplot confirms the result of the t-test =, as the difference in MPG between the automatic and manual transmissions are significantly different (p-value < 0.05).

Conclusion

Although in this data set on average manual vehicles achieve a fuel efficiency of 7.2 miles per gallon more than automatic vehicles, transmission type is not a particularly good predictor of fuel efficiency. We were able to identify that the number of cylinders and the weight of the automobile are good predictors of fuel efficiency, achieving an adjusted R squared of 0.82. If we add transmission type to this model, then the difference in fuel efficiency for a manual transmission is much smaller, just 0.18 miles per gallon for a vehicle with the same weight and number of cylinders. Therefore we conclude that number of cylinders and displacement are good predictors of fuel efficiency, but transmission type is not.

Appendix

Exploratory Analysis

```
summary(mtcars)
```

```
##      mpg      cyl      disp      hp      drat
##  Min.   :10.40   4:11   Min.   : 71.1   Min.   : 52.0   3.07   : 3
```

```
## 1st Qu.:15.43 6: 7 1st Qu.:120.8 1st Qu.: 96.5 3.92 : 3
## Median :19.20 8:14 Median :196.3 Median :123.0 2.76 : 2
## Mean :20.09 Mean :230.7 Mean :146.7 3.08 : 2
## 3rd Qu.:22.80 3rd Qu.:326.0 3rd Qu.:180.0 3.15 : 2
## Max. :33.90 Max. :472.0 Max. :335.0 3.9 : 2
## (Other):18
## wt qsec vs am gear carb
## Min. :1.513 Min. :14.50 0:18 Automatic:19 3:15 1: 7
## 1st Qu.:2.581 1st Qu.:16.89 1:14 Manual :13 4:12 2:10
## Median :3.325 Median :17.71 5: 5 3: 3
## Mean :3.217 Mean :17.85 4:10
## 3rd Qu.:3.610 3rd Qu.:18.90 6: 1
## Max. :5.424 Max. :22.90 8: 1
##
```

Regression Model Results

```
summary(best.model)
```

```
##
## Call:
## lm(formula = mpg ~ cyl + disp + hp + drat + wt + gear, data = mtcars)
##
## Residuals:
## Mazda RX4 Mazda RX4 Wag Datsun 710 Hornet 4 Drive
## 6.305e-01 -6.305e-01 2.776e-17 -1.205e-01
## Hornet Sportabout Valiant Duster 360 Merc 240D
## -9.637e-02 1.205e-01 1.665e-16 -3.053e-16
## Merc 230 Merc 280 Merc 280C Merc 450SE
## 1.305e-15 7.000e-01 -7.000e-01 -9.385e-01
## Merc 450SL Merc 450SLC Cadillac Fleetwood Lincoln Continental
## 1.643e+00 -7.044e-01 2.776e-17 -5.551e-17
## Chrysler Imperial Fiat 128 Honda Civic Toyota Corolla
## 0.000e+00 1.686e+00 -3.608e-16 -4.718e-16
## Toyota Corona Dodge Challenger AMC Javelin Camaro Z28
## -1.943e-16 -1.205e-01 9.637e-02 -2.776e-16
## Pontiac Firebird Fiat X1-9 Porsche 914-2 Lotus Europa
## 1.205e-01 -1.686e+00 1.388e-16 0.000e+00
## Ford Pantera L Ferrari Dino Maserati Bora Volvo 142E
## 8.327e-17 0.000e+00 5.551e-17 -2.498e-16
##
## Coefficients: (1 not defined because of singularities)
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 31.038 20.528 1.512 0.2051
## cyl6 -59.745 40.381 -1.480 0.2131
## cyl8 -79.839 49.838 -1.602 0.1844
## disp -1.389 1.072 -1.295 0.2649
## hp 3.259 2.443 1.334 0.2530
## drat2.93 20.944 27.242 0.769 0.4849
## drat3 -29.174 13.566 -2.151 0.0979 .
## drat3.07 -157.389 119.891 -1.313 0.2595
```

```
## drat3.08      34.305      25.051      1.369      0.2427
## drat3.15     -19.546      16.086     -1.215      0.2911
## drat3.21    -252.792     187.756     -1.346      0.2494
## drat3.23    -101.149      69.620     -1.453      0.2199
## drat3.54    -627.339     470.801     -1.332      0.2535
## drat3.62    -334.116     254.614     -1.312      0.2597
## drat3.69    -225.129     157.636     -1.428      0.2264
## drat3.7      -170.968     130.460     -1.311      0.2602
## drat3.73    -269.020     199.065     -1.351      0.2479
## drat3.77    -244.231     194.278     -1.257      0.2771
## drat3.85    -377.210     272.852     -1.382      0.2390
## drat3.9     -304.535     219.304     -1.389      0.2373
## drat3.92    -342.264     244.439     -1.400      0.2340
## drat4.08    -321.424     237.585     -1.353      0.2475
## drat4.11    -414.963     299.015     -1.388      0.2375
## drat4.22    -323.732     242.931     -1.333      0.2535
## drat4.43    -145.031     114.904     -1.262      0.2755
## drat4.93    -277.391     205.793     -1.348      0.2490
## wt           4.945        4.461      1.109      0.3298
## gear4        204.483     139.689      1.464      0.2171
## gear5          NA          NA          NA          NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.704 on 4 degrees of freedom
## Multiple R-squared:  0.9897, Adjusted R-squared:  0.9201
## F-statistic: 14.21 on 27 and 4 DF,  p-value: 0.009558
```

Boxplot

```
boxplot(mpg ~ am, data = mtcars, col = "blue", ylab = "miles per gallon")
```

