

Regression Models

Janus

Friday, November 16, 2015

Context

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”

Question

Take the mtcars data set and write up an analysis to answer their question using regression models and exploratory data analyses.

Your report must be:

Written as a PDF printout of a compiled (using knitr) R markdown document. Brief. Roughly the equivalent of 2 pages or less for the main text. Supporting figures in an appendix can be included up to 5 total pages including the 2 for the main report. The appendix can only include figures. Include a first paragraph executive summary. Upload your PDF by clicking the Upload button below the text box.

Peer Grading

The criteria that your classmates will use to evaluate and grade your work are shown below. Each criteria is binary: (1 point = criteria met acceptably; 0 points = criteria not met acceptably) Your Course Project score will be the sum of the points and will count as 40% of your final grade in the course.

Load Data

```
data(mtcars)
```

Exploratory analysis

```
#Results omitted for not having enough space and not needed.
```

```
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
```

```
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'))
```

```
#Result shown in the Appendix
summary(mtcars)
```

```
##      mpg      cyl      disp      hp      drat
## Min.   :10.40   4:11   Min.   : 71.1   Min.   : 52.0   Min.   :2.760
## 1st Qu.:15.43   6: 7   1st Qu.:120.8   1st Qu.: 96.5   1st Qu.:3.080
## Median :19.20   8:14   Median :196.3   Median :123.0   Median :3.695
## Mean   :20.09           Mean   :230.7   Mean   :146.7   Mean   :3.597
## 3rd Qu.:22.80           3rd Qu.:326.0   3rd Qu.:180.0   3rd Qu.:3.920
## Max.   :33.90           Max.   :472.0   Max.   :335.0   Max.   :4.930
##      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

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

```
## Start: AIC=76.4
## mpg ~ cyl + disp + hp + drat + wt + qsec + vs + am + gear + carb
##
##      Df Sum of Sq  RSS   AIC
## - carb  5    13.5989 134.00 69.828
## - gear  2     3.9729 124.38 73.442
## - am    1     1.1420 121.55 74.705
## - qsec  1     1.2413 121.64 74.732
## - drat  1     1.8208 122.22 74.884
## - cyl   2    10.9314 131.33 75.184
## - vs    1     3.6299 124.03 75.354
## <none>          120.40 76.403
## - disp  1     9.9672 130.37 76.948
## - wt    1    25.5541 145.96 80.562
## - hp    1    25.6715 146.07 80.588
##
```

```

## Step: AIC=69.83
## mpg ~ cyl + disp + hp + drat + wt + qsec + vs + am + gear
##
##      Df Sum of Sq  RSS   AIC
## - gear  2    5.0215 139.02 67.005
## - disp  1    0.9934 135.00 68.064
## - drat  1    1.1854 135.19 68.110
## - vs    1    3.6763 137.68 68.694
## - cyl   2   12.5642 146.57 68.696
## - qsec  1    5.2634 139.26 69.061
## <none>                134.00 69.828
## - am    1   11.9255 145.93 70.556
## - wt    1   19.7963 153.80 72.237
## - hp    1   22.7935 156.79 72.855
##
## Step: AIC=67
## mpg ~ cyl + disp + hp + drat + wt + qsec + vs + am
##
##      Df Sum of Sq  RSS   AIC
## - drat  1    0.9672 139.99 65.227
## - cyl   2   10.4247 149.45 65.319
## - disp  1    1.5483 140.57 65.359
## - vs    1    2.1829 141.21 65.503
## - qsec  1    3.6324 142.66 65.830
## <none>                139.02 67.005
## - am    1   16.5665 155.59 68.608
## - hp    1   18.1768 157.20 68.937
## - wt    1   31.1896 170.21 71.482
##
## Step: AIC=65.23
## mpg ~ cyl + disp + hp + wt + qsec + vs + am
##
##      Df Sum of Sq  RSS   AIC
## - disp  1    1.2474 141.24 63.511
## - vs    1    2.3403 142.33 63.757
## - cyl   2   12.3267 152.32 63.927
## - qsec  1    3.1000 143.09 63.928
## <none>                139.99 65.227
## - hp    1   17.7382 157.73 67.044
## - am    1   19.4660 159.46 67.393
## - wt    1   30.7151 170.71 69.574
##
## Step: AIC=63.51
## mpg ~ cyl + hp + wt + qsec + vs + am
##
##      Df Sum of Sq  RSS   AIC
## - qsec  1    2.442 143.68 62.059
## - vs    1    2.744 143.98 62.126
## - cyl   2   18.580 159.82 63.466
## <none>                141.24 63.511
## - hp    1   18.184 159.42 65.386
## - am    1   18.885 160.12 65.527
## - wt    1   39.645 180.88 69.428
##

```

```
## Step: AIC=62.06
## mpg ~ cyl + hp + wt + vs + am
##
##      Df Sum of Sq  RSS   AIC
## - vs    1      7.346 151.03 61.655
## <none>                  143.68 62.059
## - cyl    2     25.284 168.96 63.246
## - am     1     16.443 160.12 63.527
## - hp     1     36.344 180.02 67.275
## - wt     1     41.088 184.77 68.108
##
## Step: AIC=61.65
## mpg ~ cyl + hp + wt + am
##
##      Df Sum of Sq  RSS   AIC
## <none>                  151.03 61.655
## - am     1      9.752 160.78 61.657
## - cyl    2     29.265 180.29 63.323
## - hp     1     31.943 182.97 65.794
## - wt     1     46.173 197.20 68.191
```

```
#Result shown in the Appendix
summary(best.model)
```

```
##
## Call:
## lm(formula = mpg ~ cyl + hp + wt + am, data = mtcars)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.9387 -1.2560 -0.4013  1.1253  5.0513
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 33.70832    2.60489   12.940 7.73e-13 ***
## cyl6        -3.03134    1.40728   -2.154  0.04068 *
## cyl8        -2.16368    2.28425   -0.947  0.35225
## hp          -0.03211    0.01369   -2.345  0.02693 *
## wt          -2.49683    0.88559   -2.819  0.00908 **
## amManual     1.80921    1.39630    1.296  0.20646
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.41 on 26 degrees of freedom
## Multiple R-squared:  0.8659, Adjusted R-squared:  0.8401
## F-statistic: 33.57 on 5 and 26 DF,  p-value: 1.506e-10
```

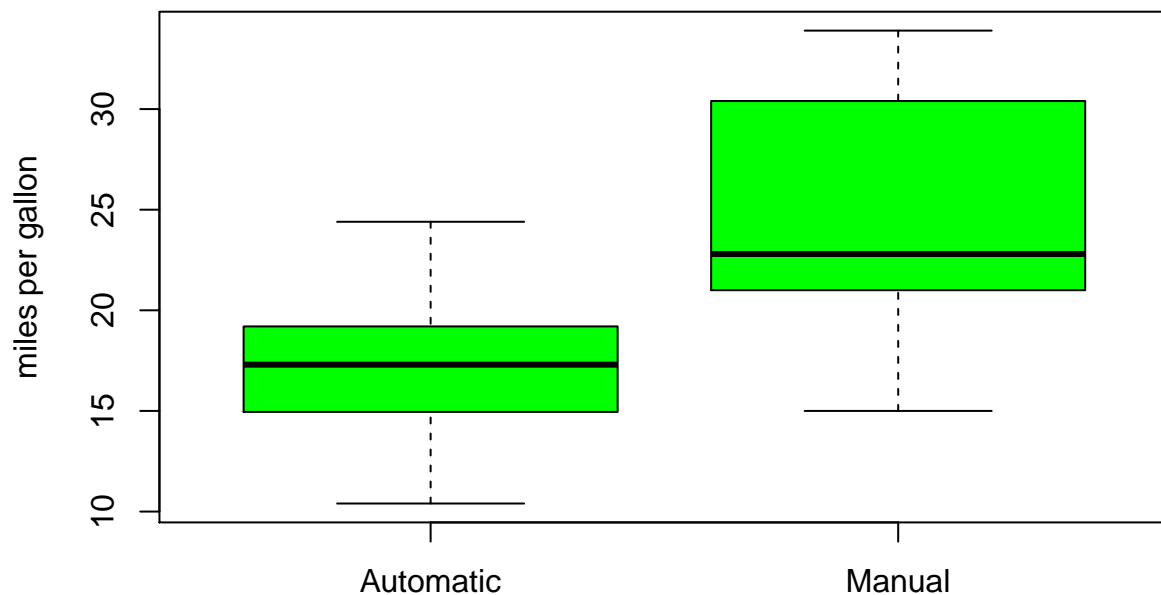
```
t.test(mpg ~ am, data = mtcars)
```

```
##
## 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 = "green", ylab = "miles per gallon")
```



*****Conclusion

According to the results, cars with a manual transmission are better for mpg than cars with an automatic transmission. The rate of change of the conditional mean mpg with respect to am is about 1.8, and we are 95% confident that this value varies between -1.06 and 4.68. There are however some limitations to this study.

Appendix

Exploratory analysis

##	mpg	cyl	displacement	hp	drat
##	Min. :10.40	4:11	Min. : 71.1	Min. : 52.0	Min. :2.760
##	1st Qu.:15.43	6: 7	1st Qu.:120.8	1st Qu.: 96.5	1st Qu.:3.080
##	Median :19.20	8:14	Median :196.3	Median :123.0	Median :3.695
##	Mean :20.09		Mean :230.7	Mean :146.7	Mean :3.597
##	3rd Qu.:22.80		3rd Qu.:326.0	3rd Qu.:180.0	3rd Qu.:3.920
##	Max. :33.90		Max. :472.0	Max. :335.0	Max. :4.930

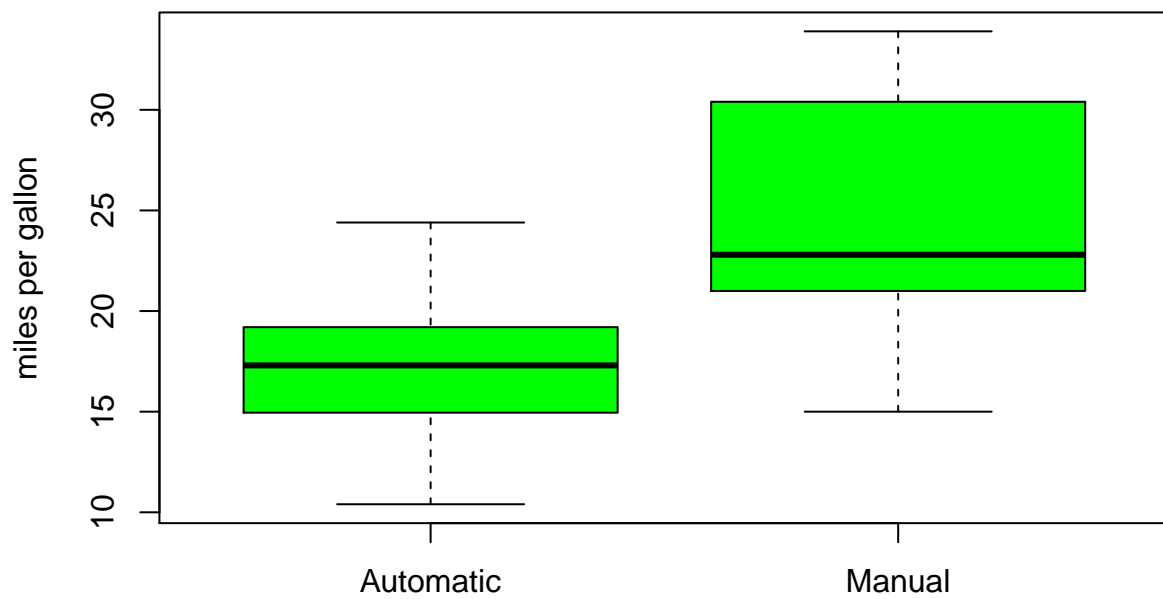
```
##           wt           qsec      vs           am      gear      carb
##  Min.      :1.513   Min.      :14.50   0:18   Automatic:19   3:15   1: 7
##  1st Qu.:2.581   1st Qu.:16.89   1:14   Manual      :13   4:12   2:10
##  Median :3.325   Median :17.71                   5: 5   3: 3
##  Mean      :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

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

```
##
## Call:
## lm(formula = mpg ~ cyl + hp + wt + am, data = mtcars)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.9387 -1.2560 -0.4013  1.1253  5.0513
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  33.70832    2.60489   12.940 7.73e-13 ***
## cyl6         -3.03134    1.40728   -2.154  0.04068 *
## cyl8         -2.16368    2.28425   -0.947  0.35225
## hp           -0.03211    0.01369   -2.345  0.02693 *
## wt           -2.49683    0.88559   -2.819  0.00908 **
## amManual      1.80921    1.39630    1.296  0.20646
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.41 on 26 degrees of freedom
## Multiple R-squared:  0.8659, Adjusted R-squared:  0.8401
## F-statistic: 33.57 on 5 and 26 DF,  p-value: 1.506e-10
```

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



```
par(mfrow=c(2, 2))  
plot(best.model)
```

