Regression Models Course Project

Nadine.solov

20 02 2022

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"

```
library(datasets)
data(mtcars)
str(mtcars)
```

```
## 'data.frame':    32 obs. of  11 variables:
##  $ mpg : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
##  $ cyl : num  6 6 4 6 8 6 8 4 4 6 ...
##  $ disp: num  160 160 108 258 360 ...
##  $ hp  : num  110 110 93 110 175 105 245 62 95 123 ...
##  $ drat: num  3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...
##  $ wt  : num  2.62 2.88 2.32 3.21 3.44 ...
##  $ qsec: num  16.5 17 18.6 19.4 17 ...
##  $ vs  : num  0 0 1 1 0 1 0 1 1 1 ...
##  $ am  : num  1 1 1 0 0 0 0 0 0 0 ...
##  $ gear: num  4 4 4 3 3 3 3 4 4 4 ...
##  $ carb: num  4 4 1 1 2 1 4 2 2 4 ...
```

```
library(car)
```

Exploratory analysis creating a multiple linear regression with multiple independent variables,

dependent variable mpg (Miles/(US) gallon), checking the summary on it

```
fitmpg<-lm(mpg~., mtcars)
summary(fitmpg)
```

```
##
## Call:
## lm(formula = mpg ~ ., data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
```

```
## -3.4506 -1.6044 -0.1196  1.2193  4.6271
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 12.30337   18.71788   0.657   0.5181
## cyl         -0.11144    1.04502  -0.107   0.9161
## disp         0.01334    0.01786   0.747   0.4635
## hp          -0.02148    0.02177  -0.987   0.3350
## drat         0.78711    1.63537   0.481   0.6353
## wt          -3.71530    1.89441  -1.961   0.0633 .
## qsec         0.82104    0.73084   1.123   0.2739
## vs           0.31776    2.10451   0.151   0.8814
## am           2.52023    2.05665   1.225   0.2340
## gear         0.65541    1.49326   0.439   0.6652
## carb        -0.19942    0.82875  -0.241   0.8122
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.65 on 21 degrees of freedom
## Multiple R-squared:  0.869,  Adjusted R-squared:  0.8066
## F-statistic: 13.93 on 10 and 21 DF,  p-value: 3.793e-07
```

The model has good indicators Adjusted R-squared: 0.8066, p-value: 3.793e-07 and shows the dependence of mpg on various factors.

We construct a diagram reflecting a linear approximation of the dependence for each pair of variables.

see the appendix (scatterplotMatrix № 1).

We select the most significant variables and use the R-function Step():

```
recuce_fitmpg<-step(fitmpg, direction = "backward")
```

```
## Start:  AIC=70.9
## mpg ~ cyl + disp + hp + drat + wt + qsec + vs + am + gear + carb
##
##        Df Sum of Sq    RSS    AIC
## - cyl   1    0.0799 147.57 68.915
## - vs    1    0.1601 147.66 68.932
## - carb  1    0.4067 147.90 68.986
## - gear  1    1.3531 148.85 69.190
## - drat  1    1.6270 149.12 69.249
## - disp  1    3.9167 151.41 69.736
## - hp    1    6.8399 154.33 70.348
## - qsec  1    8.8641 156.36 70.765
## <none>              147.49 70.898
## - am    1   10.5467 158.04 71.108
## - wt    1   27.0144 174.51 74.280
##
## Step:  AIC=68.92
## mpg ~ disp + hp + drat + wt + qsec + vs + am + gear + carb
##
##        Df Sum of Sq    RSS    AIC
## - vs    1    0.2685 147.84 66.973
## - carb  1    0.5201 148.09 67.028
```

```
## - gear  1    1.8211 149.40 67.308
## - drat  1    1.9826 149.56 67.342
## - disp  1    3.9009 151.47 67.750
## - hp    1    7.3632 154.94 68.473
## <none>              147.57 68.915
## - qsec  1   10.0933 157.67 69.032
## - am    1   11.8359 159.41 69.384
## - wt    1   27.0280 174.60 72.297
##
## Step:  AIC=66.97
## mpg ~ disp + hp + drat + wt + qsec + am + gear + carb
##
##        Df Sum of Sq    RSS    AIC
## - carb  1    0.6855 148.53 65.121
## - gear  1    2.1437 149.99 65.434
## - drat  1    2.2139 150.06 65.449
## - disp  1    3.6467 151.49 65.753
## - hp    1    7.1060 154.95 66.475
## <none>              147.84 66.973
## - am    1   11.5694 159.41 67.384
## - qsec  1   15.6830 163.53 68.200
## - wt    1   27.3799 175.22 70.410
##
## Step:  AIC=65.12
## mpg ~ disp + hp + drat + wt + qsec + am + gear
##
##        Df Sum of Sq    RSS    AIC
## - gear  1     1.565 150.09 63.457
## - drat  1     1.932 150.46 63.535
## <none>              148.53 65.121
## - disp  1    10.110 158.64 65.229
## - am    1    12.323 160.85 65.672
## - hp    1    14.826 163.35 66.166
## - qsec  1    26.408 174.94 68.358
## - wt    1    69.127 217.66 75.350
##
## 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
## - am    1    20.036 170.13 65.466
## - qsec  1    25.574 175.67 66.491
## - 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
## - qsec  1    26.470 179.91 65.255
## - am    1    32.198 185.63 66.258
## - 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
## - qsec  1    20.225 180.29 63.323
## - am    1    25.993 186.06 64.331
## - 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
## - am    1    26.178 195.46 63.908
## - qsec  1   109.034 278.32 75.217
## - wt    1   183.347 352.63 82.790
```

the most significant variables are defined by am (Transmission (0 = automatic, 1 = manual)), qsec(1/4 mile time), wt(Weight (1000 lbs)) the improved model will look like :

```
recuce_fitmpg<-lm(mpg~ wt + qsec + am, data = mtcars)
```

diagnostic charts see the appendix (plot № 2).

Conclusions: the points on the graph (1) are scattered randomly without a pattern, (2) the points are on the line - the residuals have a normal distribution.

```
summary(recuce_fitmpg)
```

```
##
## 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 ***
## 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: 1.21e-11
```

The resulting model showed that mpg depends not only on am (p-value=0.046716), but also on wt (p-value= 6.95e-06) and qsec (p-value=0.000216)

"Quantify the MPG difference between automatic and manual transmissions":

```
am0<- mtcars$am==0
recuce_fitmpg0<-lm(mpg~am0+qsec+wt, mtcars)

am1<- mtcars$am==1
recuce_fitmpg1<-lm(mpg~am1+qsec+wt, mtcars)

Confint(recuce_fitmpg0)
```

```
##              Estimate      2.5 %      97.5 %
## (Intercept) 12.553618  0.1457211 24.96151430
## am0TRUE     -2.935837 -5.8259441 -0.04573031
## qsec         1.225886  0.6345732  1.81719875
## wt          -3.916504 -5.3733342 -2.45967322
```

```
Confint(recuce_fitmpg1)
```

```
##              Estimate       2.5 %    97.5 %
## (Intercept)  9.617781 -4.63829946 23.873860
## am1TRUE      2.935837  0.04573031  5.825944
## qsec         1.225886  0.63457320  1.817199
## wt          -3.916504 -5.37333423 -2.459673
```

Transmission (1 = manual):

```
(qsec=9.617781+1.225886)
```

```
## [1] 10.84367
```

```
(wt=9.617781-3.9165)
```

```
## [1] 5.701281
```

```
(am=9.617781+2.935837)
```

```
## [1] 12.55362
```

Transmission 0 = automatic:

```
(am=12.553618-2.935837)
```

```
## [1] 9.617781
```

```
(qsec=12.553618+1.2259)
```

```
## [1] 13.77952
```

```
(wt=12.553618-3.9165)
```

```
## [1] 8.637118
```

am=1 manual am=12.5536 qsec=10.8437, wt=5.7013

am=0 automatic am =9.617781 qsec=13.7795, wt=8.6371

Conclusions:

The resulting model does not give an unambiguous answer to the question: Is an automatic or manual transmission better for MPG?

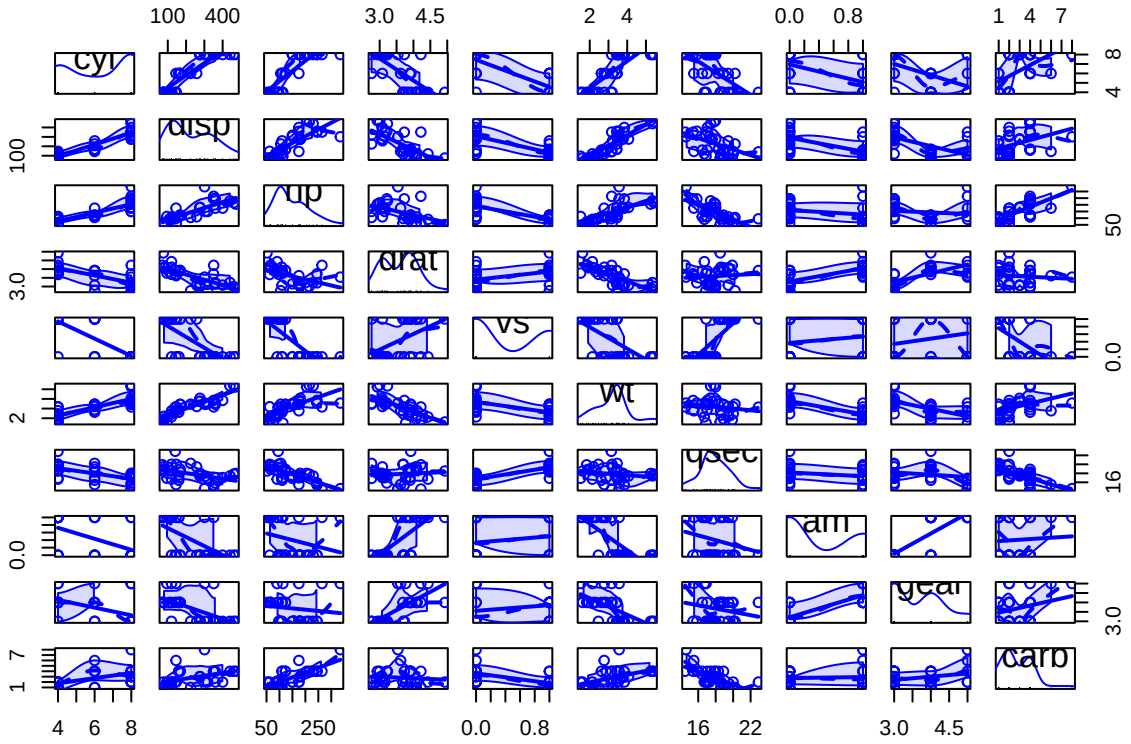
p-value=0.046716 does not give grounds to refute the null hypothesis (that there is no difference between manual transmission and automatic transmission).

MPG Transmission 0 = automatic one mile 9,6 gallon, less than Transmission 1 = manual 12,6 gallon, but qsec automatic Transmission = 13,77952 more than manual Transmission =10,8.

Appendix

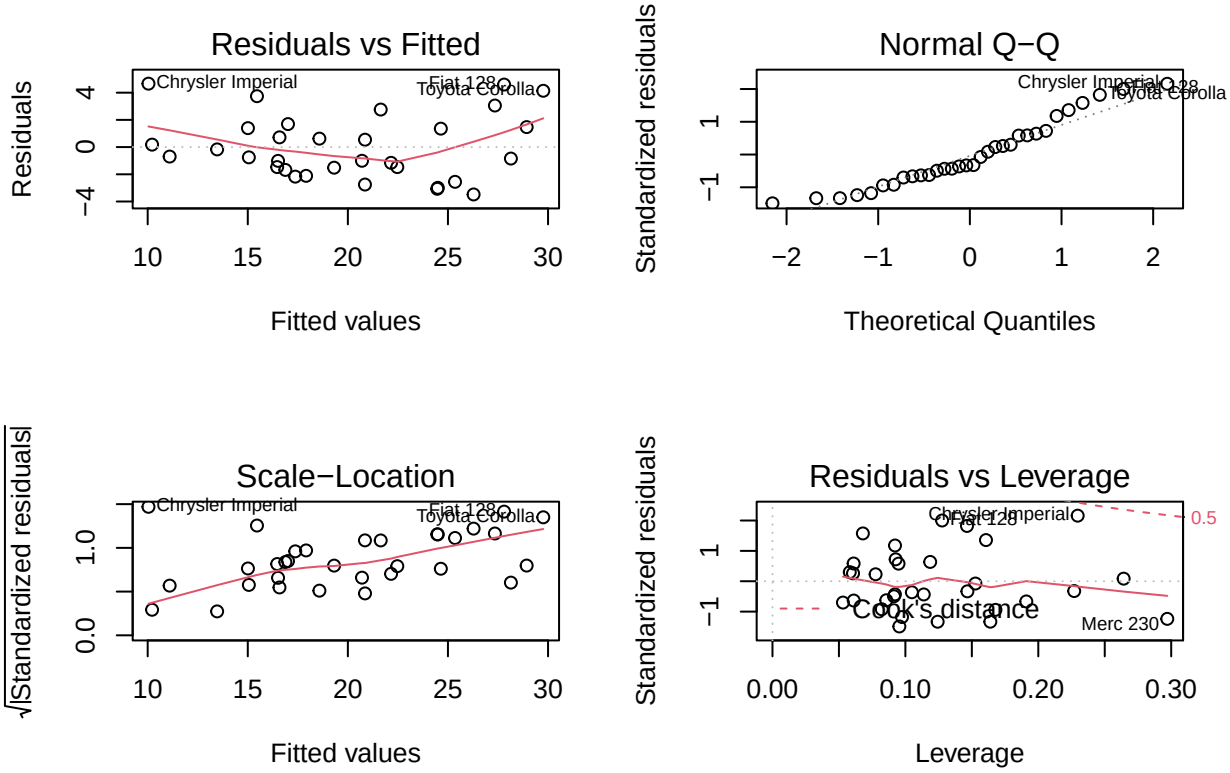
(scatterplotMatrix № 1).

```
scatterplotMatrix(~cyl+ disp + hp + drat + vs + wt + qsec + am + gear + carb, data=mtcars, diag = "boxp
```



(plot t № 2).

```
par(mfrow=(c(2,2)))
plot(recuce_fitmpg)
```


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