Comparison of cars with automatic and manual gearboxes

Pavel Naumov

05/12/2019

First glance at the Motor Trend dataset

First of all we are going to make first glance at our data set. This is the "Motor Trend Car Road Tests" dataset.The data was extracted from the 1974 Motor Trend US magazine, and comprises fuel consumption and 10 aspects of automobile design and performance for 32 automobiles (1973–74 models).

```
car<-datasets::mtcars
head(car, 10)
```

```
##                    mpg cyl  disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4         21.0   6 160.0 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag     21.0   6 160.0 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710        22.8   4 108.0  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive    21.4   6 258.0 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout 18.7   8 360.0 175 3.15 3.440 17.02  0  0    3    2
## Valiant           18.1   6 225.0 105 2.76 3.460 20.22  1  0    3    1
## Duster 360        14.3   8 360.0 245 3.21 3.570 15.84  0  0    3    4
## Merc 240D         24.4   4 146.7  62 3.69 3.190 20.00  1  0    4    2
## Merc 230          22.8   4 140.8  95 3.92 3.150 22.90  1  0    4    2
## Merc 280          19.2   6 167.6 123 3.92 3.440 18.30  1  0    4    4
```

Multivarible regression

As the first step we try to include all variables to our model and estimate their roles:

```
summary(lm(mpg ~ ., data = car))$coefficients
```

```
##                Estimate  Std. Error    t value   Pr(>|t|)
## (Intercept) 12.30337416 18.71788443  0.6573058 0.51812440
## cyl         -0.11144048  1.04502336 -0.1066392 0.91608738
## disp         0.01333524  0.01785750  0.7467585 0.46348865
## hp          -0.02148212  0.02176858 -0.9868407 0.33495531
## drat         0.78711097  1.63537307  0.4813036 0.63527790
## wt          -3.71530393  1.89441430 -1.9611887 0.06325215
## qsec         0.82104075  0.73084480  1.1234133 0.27394127
## vs           0.31776281  2.10450861  0.1509915 0.88142347
## am           2.52022689  2.05665055  1.2254035 0.23398971
## gear         0.65541302  1.49325996  0.4389142 0.66520643
## carb        -0.19941925  0.82875250 -0.2406258 0.81217871
```

As we can see a lot of variables have parameter P>0.05. As the next step we try to exclude bad variables step by step (see appendix).

As a results of backward-elimination process we have a table with only 3 columns: wt - Weight (1000 lbs); qsec - 1/4 mile time; am - Transmission (0 = automatic, 1 = manual).

```
car2<-subset(car, select = c(mpg, wt, qsec,am))
summary(lm(mpg ~ .-1, data = car2))$coefficients

##       Estimate Std. Error   t value     Pr(>|t|)
## wt   -3.185455  0.4827586 -6.598442 3.128844e-07
## qsec  1.599823  0.1021276 15.664944 1.091522e-15
## am    4.299519  1.0241147  4.198279 2.329423e-04
```

Conclusion

Base on the final model we can estimate the influence of mechanic gearbox to miles per gallon with 95% confidence.

```
fit<-summary(lm(mpg ~ . -1, data = car2))$coefficients

mechanic_gearbox<- fit[3,1]+c(-1, 1) * qt(0.975, df = (lm(mpg ~ . -1, data =
car2))$df) * fit[3, 2]
print(mechanic_gearbox)

## [1] 2.204969 6.394069
```

Mechanic gearbox increase mpg from 2.2 up to 6.4 miles per gallon with 95% confidence.

Appendix

And we can automate this backward-elimination process:

```
car2<-car
variable_name <- names(sort(summary(lm(mpg ~ ., data =
car2))$coefficients[,"Pr(>|t|)"], decreasing = TRUE)[1])
variable_P<-sort(summary(lm(mpg ~ ., data = car2))$coefficients[,"Pr(>|t|)"],
decreasing = TRUE)[1]
print(variable_P)

##       cyl
## 0.9160874

while (variable_P>0.05 && variable_name!="(Intercept)") {
car2<-car2[, -which(names(car2)==variable_name)]
print(variable_P)
variable_name<-names(sort(summary(lm(mpg ~ ., data =
car2))$coefficients[,"Pr(>|t|)"], decreasing = TRUE)[1])

variable_P<-sort(summary(lm(mpg ~ ., data = car2))$coefficients[,"Pr(>|t|)"],
```

```
decreasing = TRUE)[1]
}
```

```
##       cyl
## 0.9160874
##        vs
## 0.8432585
##      carb
## 0.7469582
##      gear
## 0.6196406
##      drat
## 0.4624012
##      disp
## 0.2989721
##        hp
## 0.2230879
```

Finally in our table we have only 3 columns wt - Weight (1000 lbs) qsec - 1/4 mile time am - Transmission (0 = automatic, 1 = manual)

```
summary(lm(mpg ~ . -1, data = car2))$coefficients
```

```
##       Estimate Std. Error   t value     Pr(>|t|)
## wt   -3.185455  0.4827586 -6.598442 3.128844e-07
## qsec  1.599823  0.1021276 15.664944 1.091522e-15
## am    4.299519  1.0241147  4.198279 2.329423e-04
```