

Atividade - ANÁLISE MULTIVARIADA 2

Charles Barros 21/10/2021

Case 2

O conjunto de dados abaixo fornece o teste de quilometragem sobre o consumo de gasolina, para veículos do ano modelo 2005 da DaimlerChrysler. A tabela completa de dados contém os dados para modelos fabricados no ano de 2005 de 21 veículos:

Base de dados

```
db <- read.csv(file = "CS2.csv", header = TRUE, sep = ";"); head(db)
```

```
##   mfr      carline car.truck cid rhp trns drv od  etw  cmp axle  n.v a.c    hc
## 1  20 300C/SRT-8          C 215 253   L5   4  2 4500   9.9 3.07 30.9   Y 0.011
## 2  20   CARAVAN          T 201 180   L4   F  2 4500   9.3 2.49 32.3   Y 0.014
## 3  20  CROSSFIRE          C 196 168   L5   R  2 3375  10.0 3.27 37.1   Y 0.001
## 4  20   DAKOTA          T 226 210   L4   R  2 4500   9.2 3.55 29.6   Y 0.012
## 5  20   DAKOTA          T 226 210   L4   4  2 5000   9.2 3.55 29.6   Y 0.011
## 6  20  DURANGO          T 348 345   L5   R  2 5250   8.6 3.55 27.2   Y 0.023
##    co co2  mpg
## 1 0.09 288 30.8
## 2 0.11 274 32.5
## 3 0.02 250 35.4
## 4 0.04 316 28.1
## 5 0.05 365 24.4
## 6 0.15 367 24.1
```

Questões

(A)

Ajuste um modelo de regressão linear múltipla usando esses dados, para estimar a milhagem de gasolina para todos os carros, usando os melhores regressores;

(B)

Calcule a estimativa e os erros padrão dos coeficientes de regressão.

(C)

Prever a quilometragem de gasolina para o primeiro veículo na tabela e explicar didaticamente o que significam as regressões.

Resposta

(A)

```
db2 <- db[, -c(1,2,8,13)]  
  
model <- lm(mpg ~ ., data = db2)
```

Para realização do modelo stepwise, foram removidos as variáveis : mfr, carline, od e a.c.

```
model1 <- lm(mpg ~ ., data = db2)
```

criando o modelo com base no conjunto de dados (db2) onde apresentam os dados necessários para aplicação da função step (ou método Stepwise).

```
step.Both <- step(model1, direction = "both", trace = 0)  
  
extractAIC(step.Both)
```

```
## [1] 7.00000 22.64398
```

Quando elaborado um modelo de regressão linear múltipla na linguagem de programação R, utilizei o comando (ou função) chamada step também conhecido como método stepwise, essa função step tem como objetivo ajudá-lo a selecionar as variáveis de entrada para o seu modelo.

Sendo assim quando utilizado o comando step, obtemos um modelo com um AIC de 22.64, vale ressaltar que a sigla AIC significa critério de informação de Akaike (AIC) é uma métrica, criada por Hirotugu Akaike, que mensura a qualidade de um

modelo estatístico. Dessa forma como demonstrado na imagem abaixo o valor de AIC do modelo foi de 22.64.

(B)

```
model.step <- lm(mpg ~ car.truck + cid + drv + etw + axle, data = db2)
```

```
summary(model.step)
```

```
##
## Call:
## lm(formula = mpg ~ car.truck + cid + drv + etw + axle, data = db2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.8939 -0.4687 -0.1156  0.4748  2.0883
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  48.338811   4.512213  10.713 3.96e-08 ***
## car.truckT   -3.452904   1.063317  -3.247 0.00584 **
## cid          -0.029287   0.004694  -6.240 2.16e-05 ***
## drvF         3.093682   1.172728   2.638 0.01948 *
## drvR         2.553456   0.938053   2.722 0.01653 *
## etw          -0.001227   0.000797  -1.540 0.14594
## axle         -1.825530   1.101510  -1.657 0.11969
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.505 on 14 degrees of freedom
## Multiple R-squared:  0.9513, Adjusted R-squared:  0.9305
## F-statistic: 45.62 on 6 and 14 DF,  p-value: 2.13e-08
```

Caso os coeficientes de todas as variáveis de nosso modelo forem igual zero então obteremos o valor de 48.33 para nossa variável dependente.

Quando utilizado o comando `summary(model.step)`, visualizamos que a variável mais significa foi a variável `cid`.

(C)

```
summary(model.step)
```

```
##
## Call:
## lm(formula = mpg ~ car.truck + cid + drv + etw + axle, data = db2)
```

```
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.8939 -0.4687 -0.1156  0.4748  2.0883
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 48.338811   4.512213  10.713 3.96e-08 ***
## car.truckT  -3.452904   1.063317  -3.247 0.00584 **
## cid         -0.029287   0.004694  -6.240 2.16e-05 ***
## drvF         3.093682   1.172728   2.638 0.01948 *
## drvR         2.553456   0.938053   2.722 0.01653 *
## etw         -0.001227   0.000797  -1.540 0.14594
## axle        -1.825530   1.101510  -1.657 0.11969
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.505 on 14 degrees of freedom
## Multiple R-squared:  0.9513, Adjusted R-squared:  0.9305
## F-statistic: 45.62 on 6 and 14 DF,  p-value: 2.13e-08

mpg_car1 = 48.338811-3.452904*0-0.029287*215+3.093682*0+2.553456*0-0.001227*4500-1

## [1] 30.91623
```

Utilizando os dados da primeira linha do banco de dados obtemos o seguinte valor de 30.91

Quando comparado com o valor oficial do banco de dados apresenta uma diferença mínima, sendo assim verificamos que nosso modelo apresenta uma boa qualidade da projeção da variável mpg.

Em meu modelo podemos visualizar quatro variáveis que impactam negativamente à medida que os valores dos coeficientes forem alterados. Já variáveis drvF e drvR impactam positivamente.

Vale destacar que as variáveis drvF e drvR são variáveis dummy, onde conseguimos visualizar na frente do nome da variável a letra referente ao dummy.