

mtcars

Grupo: Americo Freitas, Arleks dos Santos e Luciano Ozorio

2022-04-30

Analizando a correlação entre as variáveis

Em primeira análise hp está correlacionada com: mpg, cyl, disp, qsec, vs, carb. Porém, reparei diversos problemas de multicolinearidade entre as variáveis explicativas.

```
df = data.frame(mtcars)
cor(df)
```

```
##           mpg           cyl           disp           hp           drat           wt
## mpg    1.0000000 -0.8521620 -0.8475514 -0.7761684  0.68117191 -0.8676594
## cyl   -0.8521620  1.0000000  0.9020329  0.8324475 -0.69993811  0.7824958
## disp  -0.8475514  0.9020329  1.0000000  0.7909486 -0.71021393  0.8879799
## hp    -0.7761684  0.8324475  0.7909486  1.0000000 -0.44875912  0.6587479
## drat   0.6811719 -0.6999381 -0.7102139 -0.4487591  1.00000000 -0.7124406
## wt    -0.8676594  0.7824958  0.8879799  0.6587479 -0.71244065  1.0000000
## qsec   0.4186840 -0.5912421 -0.4336979 -0.7082234  0.09120476 -0.1747159
## vs     0.6640389 -0.8108118 -0.7104159 -0.7230967  0.44027846 -0.5549157
## am     0.5998324 -0.5226070 -0.5912270 -0.2432043  0.71271113 -0.6924953
## gear   0.4802848 -0.4926866 -0.5555692 -0.1257043  0.69961013 -0.5832870
## carb  -0.5509251  0.5269883  0.3949769  0.7498125 -0.09078980  0.4276059
##           qsec           vs           am           gear           carb
## mpg    0.41868403  0.6640389  0.59983243  0.4802848 -0.55092507
## cyl   -0.59124207 -0.8108118 -0.52260705 -0.4926866  0.52698829
## disp  -0.43369788 -0.7104159 -0.59122704 -0.5555692  0.39497686
## hp    -0.70822339 -0.7230967 -0.24320426 -0.1257043  0.74981247
## drat   0.09120476  0.4402785  0.71271113  0.6996101 -0.09078980
## wt    -0.17471588 -0.5549157 -0.69249526 -0.5832870  0.42760594
## qsec   1.00000000  0.7445354 -0.22986086 -0.2126822 -0.65624923
## vs     0.74453544  1.0000000  0.16834512  0.2060233 -0.56960714
## am    -0.22986086  0.1683451  1.00000000  0.7940588  0.05753435
## gear  -0.21268223  0.2060233  0.79405876  1.0000000  0.27407284
## carb  -0.65624923 -0.5696071  0.05753435  0.2740728  1.00000000
```

Modelo linear por todas as variáveis

Considerando todas as variáveis, apenas **disp** e **carb** influenciam no modelo em **0.8565**.

```
mod = lm(hp ~ ., data=df)
summary(mod)
```

```
##
## Call:
## lm(formula = hp ~ ., data = df)
##
```

```
## Residuals:
##      Min       1Q   Median       3Q      Max
## -38.681 -15.558   0.799  18.106  34.718
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   79.0484    184.5041   0.428  0.67270
## mpg          -2.0631     2.0906  -0.987  0.33496
## cyl           8.2037    10.0861   0.813  0.42513
## disp          0.4390     0.1492   2.942  0.00778 **
## drat          -4.6185    16.0829  -0.287  0.77680
## wt           -27.6600    19.2704  -1.435  0.16591
## qsec          -1.7844     7.3639  -0.242  0.81089
## vs            25.8129    19.8512   1.300  0.20758
## am             9.4863    20.7599   0.457  0.65240
## gear           7.2164    14.6160   0.494  0.62662
## carb          18.7487     7.0288   2.667  0.01441 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 25.97 on 21 degrees of freedom
## Multiple R-squared:  0.9028, Adjusted R-squared:  0.8565
## F-statistic: 19.5 on 10 and 21 DF, p-value: 1.898e-08
```

Seleção do melhor modelo pelo step

Repare que o summary da função step consegue fazer uma seleção razoável com as variáveis (**hp ~ disp + wt + carb**). A melhor combinação de variáveis sugerida chega a **0.8742** de **R-squared**.

```
mod2=step(mod, direction = "backward")
```

```
## Start:  AIC=216.97
## hp ~ mpg + cyl + disp + drat + wt + qsec + vs + am + gear + carb
##
##           Df Sum of Sq  RSS    AIC
## - qsec    1      39.6 14204 215.06
## - drat    1      55.6 14220 215.09
## - am      1     140.8 14306 215.28
## - gear    1     164.4 14329 215.34
## - cyl     1     446.2 14611 215.96
## - mpg     1     656.9 14822 216.42
## <none>          14165 216.97
## - vs      1    1140.5 15305 217.45
## - wt      1    1389.7 15554 217.96
## - carb    1    4799.2 18964 224.31
## - disp    1    5839.6 20004 226.01
##
## Step:  AIC=215.06
## hp ~ mpg + cyl + disp + drat + wt + vs + am + gear + carb
##
##           Df Sum of Sq  RSS    AIC
## - drat    1      50.6 14255 213.17
## - gear    1     184.3 14389 213.47
## - am      1     216.7 14421 213.54
## - cyl     1     565.0 14769 214.31
```

```

## - mpg 1 793.5 14998 214.80
## <none> 14204 215.06
## - vs 1 1133.8 15338 215.52
## - wt 1 2577.3 16782 218.39
## - carb 1 5653.2 19858 223.78
## - disp 1 7460.7 21665 226.57
##
## Step: AIC=213.17
## hp ~ mpg + cyl + disp + wt + vs + am + gear + carb
##
## Df Sum of Sq RSS AIC
## - gear 1 172.8 14428 211.56
## - am 1 191.5 14446 211.60
## - cyl 1 722.9 14978 212.75
## - mpg 1 848.4 15103 213.02
## <none> 14255 213.17
## - vs 1 1139.2 15394 213.63
## - wt 1 2528.3 16783 216.40
## - carb 1 5726.5 19981 221.98
## - disp 1 7415.0 21670 224.57
##
## Step: AIC=211.56
## hp ~ mpg + cyl + disp + wt + vs + am + carb
##
## Df Sum of Sq RSS AIC
## - am 1 396.7 14824 210.43
## - cyl 1 564.9 14993 210.79
## - mpg 1 808.0 15236 211.30
## <none> 14428 211.56
## - vs 1 1174.3 15602 212.06
## - wt 1 3249.9 17678 216.06
## - disp 1 8286.4 22714 224.08
## - carb 1 12801.2 27229 229.88
##
## Step: AIC=210.43
## hp ~ mpg + cyl + disp + wt + vs + carb
##
## Df Sum of Sq RSS AIC
## - cyl 1 303.2 15128 209.07
## - mpg 1 599.2 15424 209.69
## - vs 1 843.7 15668 210.20
## <none> 14824 210.43
## - wt 1 5557.0 20381 218.61
## - disp 1 9845.0 24669 224.72
## - carb 1 23251.5 38076 238.61
##
## Step: AIC=209.07
## hp ~ mpg + disp + wt + vs + carb
##
## Df Sum of Sq RSS AIC
## - vs 1 580.3 15708 208.28
## <none> 15128 209.07
## - mpg 1 1048.7 16176 209.22
## - wt 1 6327.7 21455 218.26

```

```
## - disp 1 17271.9 32400 231.44
## - carb 1 24482.5 39610 237.88
##
## Step: AIC=208.28
## hp ~ mpg + disp + wt + carb
##
##      Df Sum of Sq  RSS   AIC
## - mpg  1      856.1 16564 207.98
## <none>                15708 208.28
## - wt   1      5792.6 21501 216.32
## - disp 1     20461.9 36170 232.97
## - carb 1     26384.0 42092 237.82
##
## Step: AIC=207.98
## hp ~ disp + wt + carb
##
##      Df Sum of Sq  RSS   AIC
## <none>                16564 207.98
## - wt   1         4961 21525 214.36
## - disp 1     26844 43408 236.81
## - carb 1     36686 53250 243.34
```

```
summary(mod2)
```

```
##
## Call:
## lm(formula = hp ~ disp + wt + carb, data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -45.225 -14.235   3.879  20.621  39.785
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  53.16715   18.16036   2.928  0.00671 **
## disp         0.51663    0.07669   6.736 2.59e-07 ***
## wt        -28.59214    9.87292  -2.896  0.00725 **
## carb        23.57691    2.99391   7.875 1.41e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 24.32 on 28 degrees of freedom
## Multiple R-squared:  0.8863, Adjusted R-squared:  0.8742
## F-statistic: 72.78 on 3 and 28 DF,  p-value: 2.462e-13
```

Aparentemente, o melhor modelo linear possível é: **0,516disp; -28,592wt; 23,576carb**; Vamos seguir com este modelo, agora é ver se os pressupostos estão sendo atendidos

```
mod = lm(formula = hp ~ disp + wt + carb, data = df)
summary(mod)
```

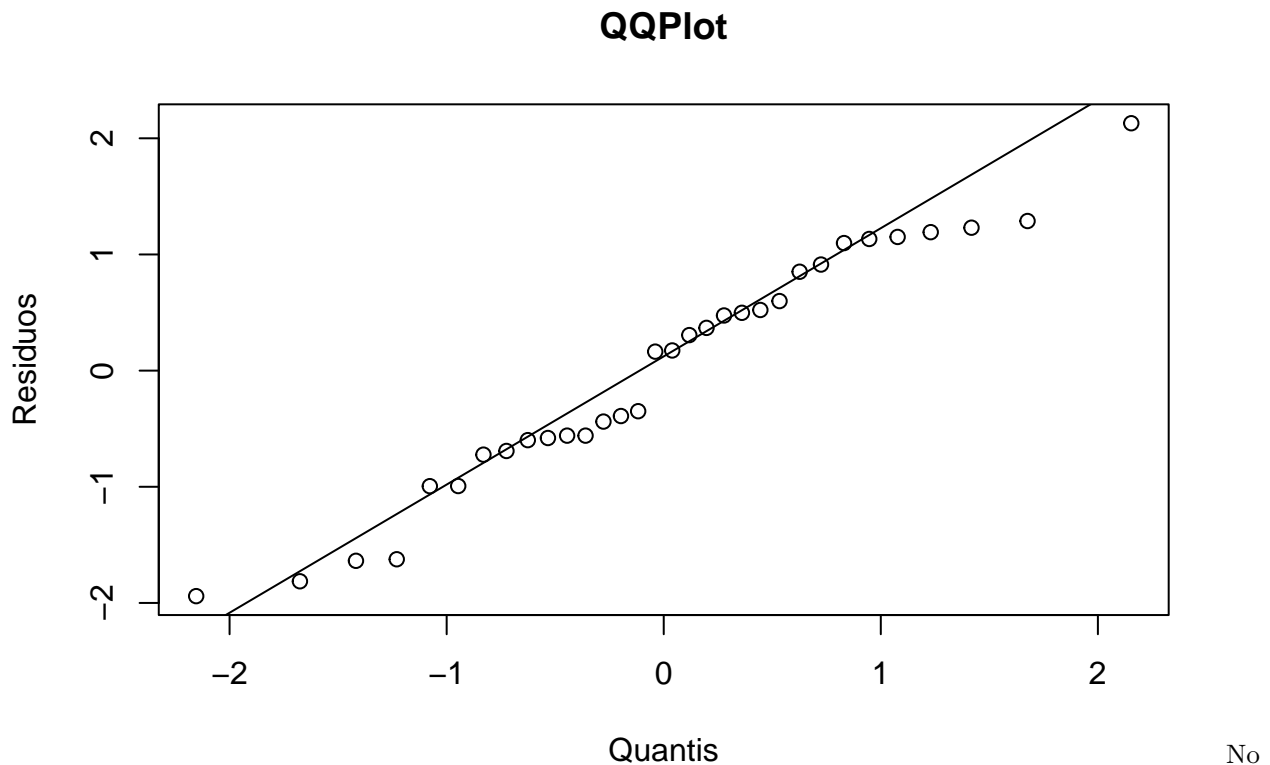
```
##
## Call:
## lm(formula = hp ~ disp + wt + carb, data = df)
##
## Residuals:
```

```
##      Min      1Q  Median      3Q      Max
## -45.225 -14.235   3.879  20.621  39.785
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  53.16715   18.16036   2.928  0.00671 **
## disp         0.51663    0.07669   6.736 2.59e-07 ***
## wt          -28.59214    9.87292  -2.896  0.00725 **
## carb         23.57691    2.99391   7.875 1.41e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 24.32 on 28 degrees of freedom
## Multiple R-squared:  0.8863, Adjusted R-squared:  0.8742
## F-statistic: 72.78 on 3 and 28 DF,  p-value: 2.462e-13
```

Teste de normalidade

Por análise visual, a normalidade dos resíduos parece ser satisfatória.

```
anares=rstandard(mod)
qqnorm(anares, ylab="Resíduos", xlab="Quantis", main="QQPlot")
qqline(anares)
```



teste formal de Normalidade (Anderson-Darling) indica 0.288, ou seja, maior que 5%, indicando que o pressuposto de normalidade foi atendido.

```
library(nortest)
ad.test(anares)
```

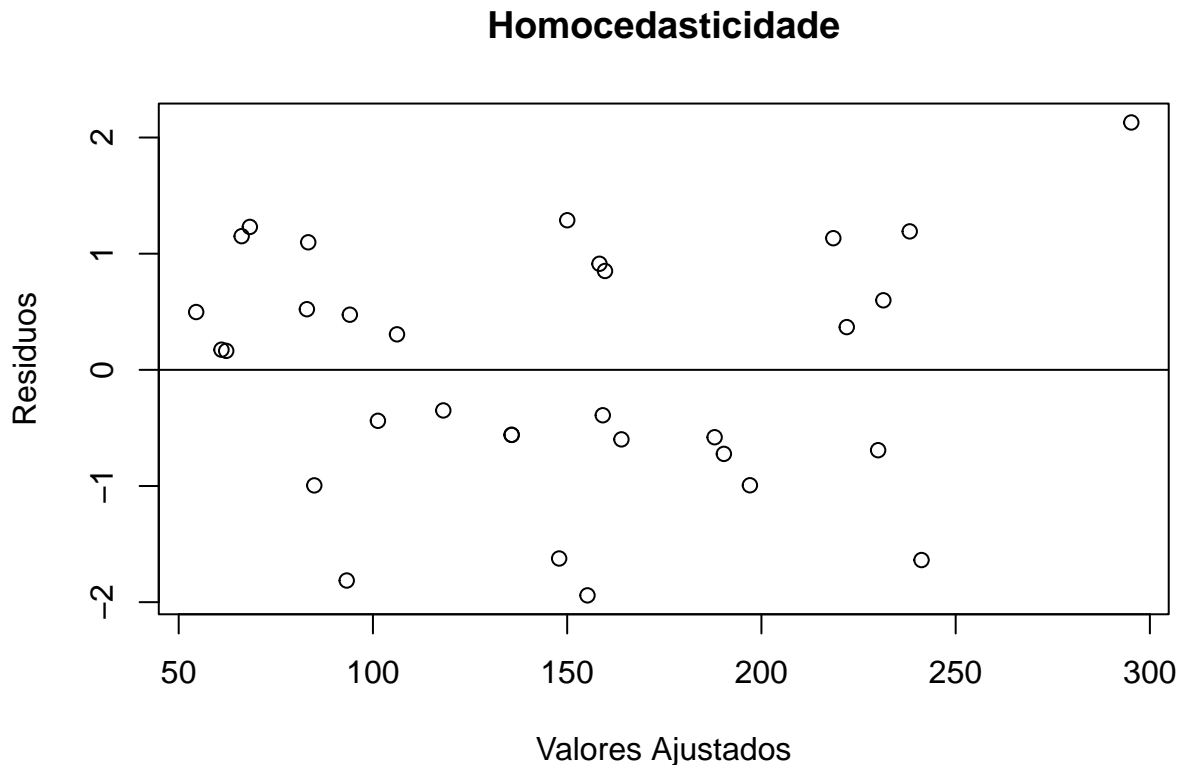
```
##
## Anderson-Darling normality test
```

```
##
## data:  anares
## A = 0.43125, p-value = 0.288
```

Teste de homocedasticidade

Por avaliação visual, a distribuição dos resíduos parecem atender ao pressuposto da homocedasticidade

```
fit=fitted.values(mod)
plot(fit, anares, ylab="Resíduos", xlab="Valores Ajustados", main="Homocedasticidade")
abline(0,0)
```



No resultado do teste formal (Breusch-Pagan) 0,09634 indica que os resíduos atendem ao pressuposto de homocedasticidade por ser maior que 5%.

```
library(zoo)
```

```
##
## Attaching package: 'zoo'
## The following objects are masked from 'package:base':
##
##   as.Date, as.Date.numeric
```

```
library(lmtest)
bptest(mod)
```

```
##
## studentized Breusch-Pagan test
##
## data:  mod
## BP = 6.3366, df = 3, p-value = 0.09634
```

Teste de autocorrelação

No resultado do teste formal de autocorrelação (Durbin-Watson) 0,3739 maior que 5% indica que o pressuposto de autocorrelação foi atendido.

```
dwtest(mod)

##
## Durbin-Watson test
##
## data:  mod
## DW = 1.9832, p-value = 0.3739
## alternative hypothesis: true autocorrelation is greater than 0
```

Até aqui todos os pressupostos foram atendidos, e parece que chegamos no melhor ajuste sem ser necessário transformações nos dados.

Previsões e intervalos de confiança

A previsão de hp para os parametros passados com base no modelo, foi **142,5679**. E foi estabelecido um intervalo de confiança entre **127,453** e **157,6827**.

```
pred_in = data.frame(displacement = 159.0, weight = 2.220, carb = 3)
predict(mod, pred_in, interval="confidence")
```

```
##           fit          lwr          upr
## 1 142.5679 127.453 157.6827
```