Estatística Aplicada II

Primeira Lista de Exercícios

SAMUEL KOJICOVSKI

Com a base de dados “ceo” obter os seguintes resultados com o auxílio do “R”

a) Elaborar a regressão linear preliminar com as seguintes variáveis:

Variável dependente: salary;

Variáveis explicativas: age, college, grad, comten, ceoten, sales, profits, mktval, profmarg

```
> summary(ceo)
     salary           age           college           grad           comten          ceoten      
 Min.   : 100.0   Min.   :33.00   Min.   :0.0000   Min.   :0.0000   Min.   : 2.0   Min.   : 0.000  
 1st Qu.: 471.0   1st Qu.:52.00   1st Qu.:1.0000   1st Qu.:0.0000   1st Qu.:12.0   1st Qu.: 3.000  
 Median : 707.0   Median :57.00   Median :1.0000   Median :1.0000   Median :23.0   Median : 6.000  
 Mean   : 865.9   Mean   :56.43   Mean   :0.9718   Mean   :0.5311   Mean   :22.5   Mean   : 7.955  
 3rd Qu.:1119.0   3rd Qu.:62.00   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:33.0   3rd Qu.:11.000  
 Max.   :5299.0   Max.   :86.00   Max.   :1.0000   Max.   :1.0000   Max.   :58.0   Max.   :37.000  
```

```
     ceoten           sales           profits           mktval          profmarg       
 Min.   : 0.000   Min.   :   29   Min.   :-463.0   Min.   :  387   Min.   :-203.077  
 1st Qu.: 3.000   1st Qu.:  561   1st Qu.:  34.0   1st Qu.:  644   1st Qu.:   4.231  
 Median : 6.000   Median : 1400   Median :  63.0   Median : 1200   Median :   6.834  
 Mean   : 7.955   Mean   : 3529   Mean   : 207.8   Mean   : 3600   Mean   :   6.420  
 3rd Qu.:11.000   3rd Qu.: 3500   3rd Qu.: 208.0   3rd Qu.: 3500   3rd Qu.:  10.947  
 Max.   :37.000   Max.   :51300   Max.   :2700.0   Max.   :45400   Max.   :  47.458  
```

```
> result <- lm(salary ~ age + college + grad + comten + ceoten + sales + profits
+              + mktval + profmarg, data = ceo)
> summary(result)

Call:
lm(formula = salary ~ age + college + grad + comten + ceoten + 
    sales + profits + mktval + profmarg, data = ceo)

Residuals:
    Min      1Q  Median      3Q     Max 
-1108.9  -272.7  -104.8   212.1  4485.7 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)  
(Intercept) 711.48819  401.30075   1.773   0.0781 .
age           3.17751    5.67010   0.560   0.5760  
college    -132.53517  250.59622  -0.529   0.5976  
grad        -56.05080   84.82088  -0.661   0.5096  
comten       -5.27596    3.91034  -1.349   0.1791  
ceoten       13.75820    6.17652   2.228   0.0273 *
sales         0.01606    0.01133   1.417   0.1582  
profits       0.10527    0.28315   0.372   0.7105  
mktval        0.02115    0.01606   1.316   0.1898  
profmarg     -1.83252    2.33473  -0.785   0.4336  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 534.4 on 167 degrees of freedom
Multiple R-squared:  0.2153,    Adjusted R-squared:  0.173 
F-statistic: 5.091 on 9 and 167 DF,  p-value: 4.383e-06
```

b) Testar outliers e deletar se necessário (essa etapa é opcional);

```
> outlierTest(result1)
    rstudent unadjusted p-value Bonferroni p
103 11.98232         4.3018e-24   7.6141e-22
74   3.93273         1.2427e-04   2.1996e-02
```

```
> summary(result1)

Call:
lm(formula = salary ~ age + college + grad + comten + ceoten +
    sales + profits + mktval + profmarg + ceoten2 + salessqrt +
    mktvalsqrt + profmarg_prof2, data = ceo)

Residuals:
     Min       1Q   Median       3Q      Max
-1138.13  -208.92   -58.19   211.33   971.11

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)     4.870e+02  2.699e+02   1.804 0.073044 .
age             4.160e+00  3.645e+00   1.141 0.255513
college        -1.721e+02  1.589e+02  -1.083 0.280319
grad           -5.363e+00  5.490e+01  -0.098 0.922295
comten         -9.026e+00  2.525e+00  -3.575 0.000463 ***
ceoten          3.161e+01  9.795e+00   3.227 0.001516 **
sales          -9.956e-03  1.560e-02  -0.638 0.524361
profits         2.250e-01  1.852e-01   1.215 0.226105
mktval         -1.153e-02  1.794e-02  -0.643 0.521184
profmarg       -1.491e+01  3.839e+00  -3.883 0.000150 ***
ceoten2        -7.259e-01  3.273e-01  -2.217 0.027995 *
salessqrt       2.115e+00  3.149e+00   0.672 0.502816
mktvalsqrt      7.558e+00  3.423e+00   2.208 0.028672 *
profmarg_prof2 -8.242e-02  2.215e-02  -3.721 0.000274 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 335.2 on 161 degrees of freedom
Multiple R-squared:  0.5133,    Adjusted R-squared:  0.474
F-statistic: 13.06 on 13 and 161 DF,  p-value: < 2.2e-16
```

c) Testar a especificação do modelo e alterar se for o caso;

Reset test

```
RESET = 1.1618, df1 = 26, df2 = 130, p-value = 0.2853
```

```
> qf(.95, df1=54, df2=130)
[1] 1.436365
```

d) Testar autocorrelação e corrigir com HAC se for o caso;

Não há autocorrelação.

```
DW = 1.9532, p-value = 0.3715
alternative hypothesis: true autocorrelation is greater than 0
```

```
> coeftest(result1, vcov. = vcovHAC)

t test of coefficients:

                  Estimate  Std. Error t value  Pr(>|t|)
(Intercept)     4.8068e+02  3.0963e+02  1.5524 0.1225928
age             3.9043e+00  3.7038e+00  1.0541 0.2934504
college        -1.6755e+02  2.7347e+02 -0.6127 0.5409860
grad           -8.3304e+00  4.6617e+01 -0.1787 0.8584077
comten         -8.5417e+00  2.1760e+00 -3.9255 0.0001296 ***
ceoten          3.2879e+01  9.4785e+00  3.4688 0.0006759 ***
sales          -5.6928e-03  1.7462e-02 -0.3260 0.7448594
profits         2.1804e-01  2.1812e-01  0.9997 0.3190169
mktval         -1.5640e-02  1.8700e-02 -0.8363 0.4042409
profmarg       -1.4876e+01  5.2841e+00 -2.8152 0.0055042 **
ceoten2        -7.6330e-01  3.4985e-01 -2.1818 0.0306226 *
salessqrt       1.0672e+00  3.9724e+00  0.2686 0.7885557
mktvalsqrt      8.6423e+00  4.3394e+00  1.9916 0.0481601 *
profmarg_prof2 -8.2542e-02  2.8629e-02 -2.8831 0.0044934 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

e) Testar heterocedasticidade e corrigir se for o caso, com HC1;

```
BP = 67.86, df = 13, p-value = 1.984e-09
```

Valor do teste maior que valor tabelado, ou seja, rejeito H0, variância não constante.

```
> chisup <- qchisq(.95, df = 13)
> chisup
[1] 22.36203
```

Correção com HC1

```
> coeftest(result1, vcov=vcovHC(result1, type="HC1"))

t test of coefficients:

                  Estimate  Std. Error t value  Pr(>|t|)
(Intercept)     4.8068e+02  3.0754e+02  1.5630 0.1200847
age             3.9043e+00  3.6724e+00  1.0631 0.2893582
college        -1.6755e+02  2.7805e+02 -0.6026 0.5476602
grad           -8.3304e+00  5.3994e+01 -0.1543 0.8775857
comten         -8.5417e+00  2.1784e+00 -3.9212 0.0001317 ***
ceoten          3.2879e+01  9.5452e+00  3.4446 0.0007348 ***
sales          -5.6928e-03  1.6268e-02 -0.3499 0.7268515
profits         2.1804e-01  2.1935e-01  0.9940 0.3217470
mktval         -1.5640e-02  1.8557e-02 -0.8428 0.4006281
profmarg       -1.4876e+01  5.2441e+00 -2.8367 0.0051642 **
ceoten2        -7.6330e-01  3.5566e-01 -2.1461 0.0334073 *
salessqrt       1.0672e+00  3.9722e+00  0.2687 0.7885430
mktvalsqrt      8.6423e+00  4.2378e+00  2.0393 0.0431028 *
profmarg_prof2 -8.2542e-02  2.8405e-02 -2.9059 0.0041949 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

f) Fazer regressão stepwise e fazer a regressão do melhor modelo estimador HC1 ou HAC

```
Step:  AIC=2005.77
salary ~ comten + ceoten + profmarg + ceoten2 + mktvalsqrt +
    profmarg_prof2

               Df Sum of Sq      RSS    AIC
<none>                      18310289 2005.8
- ceoten2       1    631425 18941714 2006.4
- comten        1   1015585 19325874 2009.8
- ceoten        1   1364405 19674694 2012.8
- profmarg_prof2  1   2534278 20844567 2022.7
- profmarg      1   2665659 20975948 2023.7
- mktvalsqrt    1  14739343 33049632 2101.0

Call:
lm(formula = salary ~ comten + ceoten + profmarg + ceoten2 +
    mktvalsqrt + profmarg_prof2, data = ceo)

Coefficients:
   (Intercept)         comten         ceoten       profmarg        ceoten2      mktvalsqrt  profmarg_prof2
      523.9327        -6.8651        33.9594       -14.7368        -0.7754         8.3353         -0.0835
```

```
> summary(result_final)

Call:
lm(formula = salary ~ comten + ceoten + profmarg + ceoten2 +
    mktvalsqrt + profmarg_prof2, data = ceo)

Residuals:
     Min       1Q   Median       3Q      Max
-1207.88  -200.35   -57.49   229.62  1056.46

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    523.93266   80.92112   6.475 1.07e-09 ***
comten          -6.86513    2.28320  -3.007 0.003059 **
ceoten          33.95941    9.74411   3.485 0.000632 ***
profmarg       -14.73678    3.02520  -4.871 2.61e-06 ***
ceoten2         -0.77537    0.32704  -2.371 0.018915 *
mktvalsqrt       8.33533    0.72768  11.455  < 2e-16 ***
profmarg_prof2  -0.08350    0.01758  -4.750 4.44e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 335.2 on 163 degrees of freedom
Multiple R-squared:  0.4973,    Adjusted R-squared:  0.4788
F-statistic: 26.87 on 6 and 163 DF,  p-value: < 2.2e-16
```

Recalculando heterocedasticidade

```
BP = 24.551, df = 6, p-value = 0.0004134

> chisup <- qchisq(.95, df = 24.551)
> chisup
[1] 37.09748
```

g) Obter o AIC, BIC e AICc do melhor modelo selecionado.

Analisando todos os modelos, foi selecionado (**result1**) o que se retirou os outliers, com um R^2 de 51.5%, sendo melhor que o indicado pelo stepwise.

```
> AICc (result1)
[1] 2479.416
> AIC (result1)
[1] 2476.299
> BIC (result1)
[1] 2523.336
> AICc (result1)
[1] 2479.416
```

```
> model_performance(result1)
# Indices of model performance

AIC      |      BIC |    R2 | R2 (adj.) |    RMSE |   Sigma
-----------------------------------------------------------
2476.299 | 2523.336 | 0.515 |     0.474 | 322.486 | 336.646
> model_performance(result_final)
# Indices of model performance

AIC      |      BIC |    R2 | R2 (adj.) |    RMSE |   Sigma
-----------------------------------------------------------
2468.259 | 2493.345 | 0.497 |     0.479 | 328.188 | 335.161
```

h) Estimar os Intervalos de confiança dos parâmetros.

```
> confint(result1)
                    2.5 %        97.5 %
(Intercept)     -56.96253331 1018.31377965
age              -3.38371643   11.19239999
college        -483.01365799  147.92226697
grad           -118.46176614  101.80104374
comten          -13.59409503   -3.48939383
ceoten           13.20750702   52.55050191
sales            -0.03743843    0.02605290
profits          -0.14970825    0.58579548
mktval           -0.05175036    0.02047105
profmarg        -22.50020914   -7.25113889
ceoten2          -1.41800544   -0.10859416
salessqrt        -5.39761628    7.53199346
mktvalsqrt        1.67505683   15.60956965
profmarg_prof2   -0.12652774   -0.03855532
```