

```
options ls=80 ps=60 nodate;

libname reg "/home/u36587463";

proc import

datafile= "/home/u36587463/Regressão/Dados_exemplo9_2_comp.xlsx" out=
reg.exemplo9_2_const DBMS=xlsx REPLACE;

sheet="Dados_exemplo9_2_const";

GETNAMES= YES;

run;
```

```
data reg.exemplo9_2_const2;

    set reg.exemplo9_2_const;

    lny = log(y);

    if x7 = 2 then x7a = 1 ; else x7a = 0;

    if x7 = 3 then x7b = 1 ; else x7b = 0;

    Label x7a = 'Uso de alcool - moderado'

    x7b = 'Uso de alcool - intenso'

    Y = "tempo de sobrevivencia após cirurgia de fígado"

    X1 = "Índice de coagulação"

    X2 = "Prognóstico (Índice)"

    X3 = "Teste de função da enzima"

    X4 = "Teste de função do fígado"

    X5 = "Idade (em anos)"

    X6 = "Gênero"

    X7 = "Uso de álcool";

run;
```

```
proc format;

value X6f 0 ="masculino" 1 = "feminino";

value Xwf 1 ="nenhum" 2 = "moderado" 3 = "frequente";

run;
```

```
/* Critérios de seleção */
```

```
symbol v=dot h=1 c=blue;  
proc reg data=reg.exemplo9_2_const2 plots=(criteria(label));  
model lny=x1 x2 x3 x4 x5 x6 x7a x7b  
/selection=rsquare sse adjrsq cp aic bic;  
run;
```

```
/* Backward */
```

```
proc reg data=reg.exemplo9_2_const2;  
model lny=x1 x2 x3 x4 x5 x6 x7a x7b/selection=Backward;  
run;
```

```
/* Esse critério selecionou as variáveis x1, x2, x3 e x7a */
```

```
/* Forward */
```

```
proc reg data=reg.exemplo9_2_const2;  
model lny=x1 x2 x3 x4 x5 x6 x7a x7b/selection=Forward;  
run;
```

```
/* Esse critério selecionou as variáveis x1, x2, x3, x5, x6, x7a e x7b */
```

```
/* Stepwise */
```

```
proc reg data=reg.exemplo9_2_const2;  
model lny=x1 x2 x3 x4 x5 x6 x7a x7b/selection=Stepwise;  
run;
```

```
/* Esse critério selecionou as variáveis x1, x2, x3, x6 e x7a */
```

```
proc reg data=reg.exemplo9_2_const2
```

```
plots=(diagnostics(stats=all)CRITERIONPANEL
```

```
RStudentByLeverage(label)
```

```
CooksD(label) criteria rsquare cp
```

```
DFFITS(label) DFBETAS (label)
```

```
ObservedByPredicted(label));
```

```
id id;
```

```
model lny = x1 x2 x3 x6 x7a/influence p r;
```

```
run;
```

```
/* Verificando o modelo sem a variável x6, que aparenta ser a menos influente no modelo */
```

```
proc reg data=reg.exemplo9_2_const2
```

```
plots=(diagnostics(stats=all)CRITERIONPANEL
```

```
RStudentByLeverage(label)
```

```
CooksD(label) criteria rsquare cp
```

```
DFFITS(label) DFBETAS (label)
```

```
ObservedByPredicted(label));
```

```
id id;
```

```
model lny = x1 x2 x3 x7a/influence p r;
```

```
run;
```

```
/* A diferença é bem pequena, porém por ser uma variável tão acessível e fácil de coletar, irei manter no modelo */
```

```
/* Verificando possíveis multicolinearidades */
```

```
/* Olhando as correlações gerais dos dados */
```

```
proc corr data=reg.exemplo9_2_const2 plot=matrix;
```

```
/* Verificando as correlações das variáveis inseridas no modelo */
```

```
proc corr data=reg.exemplo9_2_const2 plot=matrix;
```

```
var lny x1 x2 x3 x6 x7a;
```

```
run;
```

```
/* Pelas correlações, não aparenta haver forte colinearidade entre as variáveis selecionadas */
```

```
/* Modelo de validação */
```

```
/* Dados */
```

```
proc import
```

```
datafile= "/home/u36587463/Regressão/Dados_exemplo9_2_comp.xlsx" out=  
reg.exemplo9_2_valid DBMS=xlsx REPLACE;
```

```
sheet="Dados_exemplo9_2_valid";
```

```
GETNAMES= YES;
```

```
run;
```

```
data reg.exemplo9_2_valid;
```

```
set reg.exemplo9_2_valid;
```

```
lny = log(y);
```

```
if x7 = 2 then x7a = 1 ; else x7a = 0;
```

```
if x7 = 3 then x7b = 1 ; else x7b = 0;
```

```
Label x7a = 'Uso de alcool - moderado'
```

```
x7b = 'Uso de alcool - intenso'
```

```
Y = "tempo de sobrevivencia após cirurgia de fígado"
```

```
X1 = "Indice de coagulação"
```

```
X2 = "Prognóstico (Indice)"
```

```
X3 = "Teste de função da enzima"
```

```
X4 = "Teste de função do fígado"
```

```
X5 = "Idade (em anos)"
```

```
X6 = "Gênero"
```

```
X7 = "Uso de álcool";
```

```
run;
```

```

proc format;
value X6f 0 ="masculino" 1 = "feminino";
value Xwf 1 ="nenhum" 2 = "moderado" 3 = "frequente";
run;

/* Modelo com as variáveis selecionadas no conjunto de dados de validação */
symbol v=dot h=1 c=blue;
proc reg data=reg.exemplo9_2_valid plots=(criteria(label));
model lny = x1 x2 x3 x6 x7a /selection=rsquare sse adjrsq cp aic bic;
run;

/* Testes de validação */

proc reg data=reg.exemplo9_2_valid
plots=(diagnostics(stats=all)CRITERIONPANEL
RStudentByLeverage(label)
CooksD(label) criteria rsquare cp
DFFITS(label) DFBETAS (label)
ObservedByPredicted(label));
id id;
model lny = x1 x2 x3 x6 x7a/p r influence;
run;

/* Convém analisar com calma cada um dos indicadores, porém nada salta aos olhos indicando
grandes falhas no modelo */

/* Portanto, aparenta ser válido o modelo */

/* Verificando a esperança de ln(y); comparando com o ln(y) real, calculando as diferenças do
esperado para o observado; */

/* E verificando o ganho com a variável x6 */

data reg.exemplo9_2_valid;

```

```

set reg.exemplo9_2_valid;

ElNy = 3.70373 + x1*0.11108 + x2*0.01588 + x3*0.01504 + x6*0.05997 + x7a*0.17084;

Diff = ElNy-lNy;

ElNy2 = 3.70373 + x1*0.11108 + x2*0.01588 + x3*0.01504 + x7a*0.17084;

Diff2 = ElNy2-lNy;

run;

```

```

PROC MEANS DATA=reg.exemplo9_2_valid SUM;

VAR Diff Diff2 ;

RUN;

```

```

data reg.exemplo9_2_const2;

set reg.exemplo9_2_valid;

ElNy = 3.70373 + x1*0.11108 + x2*0.01588 + x3*0.01504 + x6*0.05997 + x7a*0.17084;

Diff = ElNy-lNy;

ElNy2 = 3.70373 + x1*0.11108 + x2*0.01588 + x3*0.01504 + x7a*0.17084;

Diff2 = ElNy2-lNy;

run;

```

```

PROC MEANS DATA=reg.exemplo9_2_const2 SUM;

VAR Diff Diff2 ;

RUN;

```

/* Apesar do baixo ganho de explicação, a variável x6 (gênero) não atrapalha o modelo; inclusive o melhora um pouco. */

/* Apesar de não ser muito influente, como é uma variável extremamente simples de se coletar, acho por bem manter no modelo. */

/* Calculando o MSPR */

```

PROC REG DATA=reg.exemplo9_2_const2;

MODEL lNy = x1 x2 x3 x6 x7a;

```

```
OUTPUT OUT=predicted PREDICTED=previsao;  
RUN;
```

```
PROC SQL;  
SELECT MEAN((lny-previsao)**2) AS mspr  
INTO :mspr_value  
FROM predicted;  
QUIT;
```

```
%PUT MSPR: &mspr_value;
```

```
/* Com o modelo validado, produzindo previsões próximas tanto para o modelo de treino quanto o  
de teste, com um MSPR aceitável */
```

```
/* Temos indícios para acreditar que este é um bom modelo. */
```

```
/* Importante destacar que logo ao início, a variável resposta fora transformada em logarítmo.  
Portanto, a resposta
```

```
final "comercial" deve ser aplicada a exponencial, para apresentar os resultados ao "chefe"!*/
```