ME613 - Análise de Regressão

Parte 11

Benilton S Carvalho - 1S2020

Critérios para Seleção de Modelos

Introdução

Fases na construção de um modelo:

- Coleta e preparação dos dados.
- Redução do número de variáveis preditoras.
- Refinamento e seleção de modelo.
- Validação do modelo.

Introdução

Se tivermos $p-1$ variáveis preditoras, podemos construir 2^{p-1} modelos diferentes.

Mesmo se considerarmos todos esses modelos (computacionalmente intenso), precisaríamos de algum critério para selecionar entre eles.

Métodos para seleção de modelos/variáveis foram desenvolvidos para identificar um subgrupo de variáveis que são “boas” para o modelo, segundo algum critério.

Há vários critérios desenvolvidos na literatura. Neste curso, focaremos em seis.

R_p^2

Para o critério R_p^2, a idéia é utilizar o coeficiente de determinação, R^2 para identificar subgrupos das variáveis preditoras que, quando incluídas no modelo, produzem um alto valor para R^2.

R_p^2 indica que temos p parâmetros no modelo, isto é, $p-1$ variáveis preditoras incluídas no modelo.

$$R_p^2 = 1 - \frac{SQE_p}{SQT}$$

O objetivo deste critério não é maximização: R_p^2 sempre irá aumentar conforme mais variáveis preditoras são incluídas no modelo. A idéia é comparar os diversos R_p^2's e verificar se adicionar mais variáveis ainda traz um aumento.

Exemplo: Cirurgias

Y: tempo de sobrevivência

X_1: blood clotting score

X_2: índice de prognóstico

X_3: teste de função enzimática

X_4: teste de função do fígado

X_5: idade (anos)

X_6: gênero (0=masculino, 1=feminino)

X_7: uso de álcool (1 = moderado, 0 = nenhum ou severo)

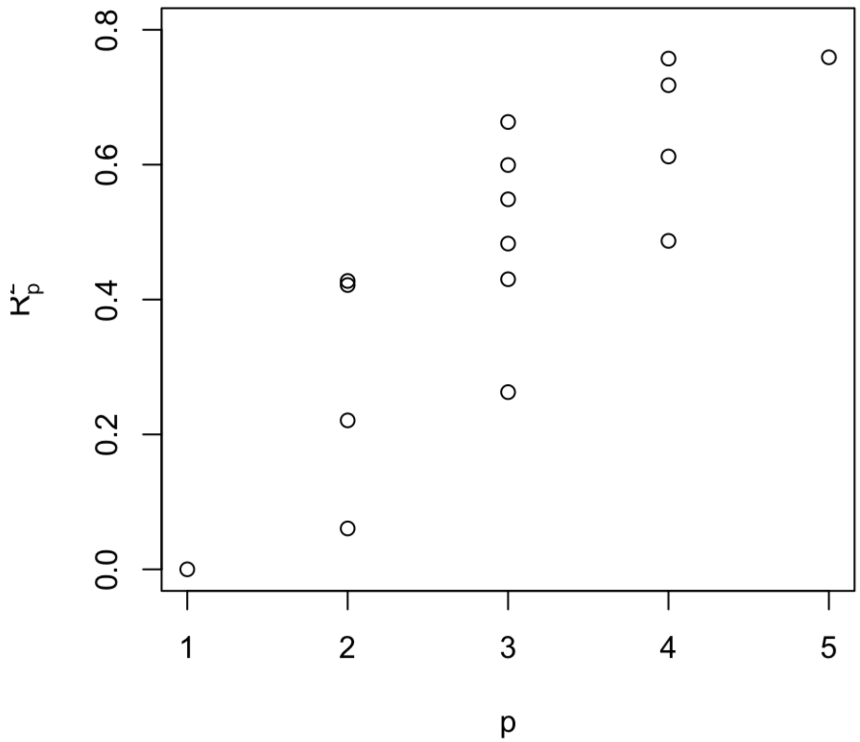
X_8: uso de álcool (1 = severo, 0 = nenhum ou moderado)

Exemplo

Considerando X_1, X_2, X_3 e X_4, temos $2^4 = 16$ modelos possíveis.

Variáveis no modelo	p	R_p^2	Variáveis no modelo	p	R_p^2
nenhuma	1	0	$X_2\ X_3$	3	0.663
X_1	2	0.061	$X_2\ X_4$	3	0.483
X_2	2	0.221	$X_3\ X_4$	3	0.599
X_3	2	0.428	$X_1\ X_2\ X_3$	4	0.757
X_4	2	0.422	$X_1\ X_2\ X_4$	4	0.487
$X_1\ X_2$	3	0.263	$X_1\ X_3\ X_4$	4	0.612
$X_1\ X_3$	3	0.549	$X_2\ X_3\ X_4$	4	0.718
$X_1\ X_4$	3	0.43	$X_1\ X_2\ X_3\ X_4$	5	0.759

Exemplo



$R^2_{a,p}$

Como R^2_p não leva em conta o número de parâmetros no modelo e sempre aumenta conforme temos mais variáveis incluídas, uma alternativa é usar:

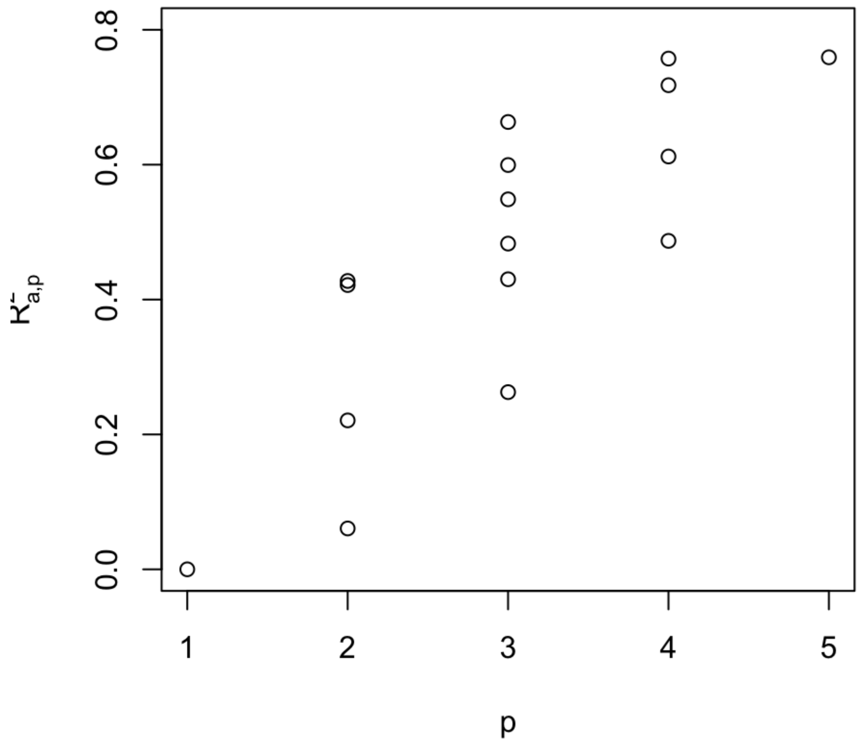
$$R^2_{a,p} = 1 - \left(\frac{n-1}{n-p}\right)\frac{SQE_p}{SQT} = 1 - \frac{QME_p}{SQT/(n-1)}$$

$R^2_{a,p}$ aumenta se e somente se QME_p diminui.

Exemplo

Variáveis no modelo	p	$R^2_{a,p}$	Variáveis no modelo	p	$R^2_{a,p}$
nenhuma	1	0	X_2 X_3	3	0.65
X_1	2	0.043	X_2 X_4	3	0.463
X_2	2	0.206	X_3 X_4	3	0.584
X_3	2	0.417	X_1 X_2 X_3	4	0.743
X_4	2	0.41	X_1 X_2 X_4	4	0.456
X_1 X_2	3	0.234	X_1 X_3 X_4	4	0.589
X_1 X_3	3	0.531	X_2 X_3 X_4	4	0.701
X_1 X_4	3	0.408	X_1 X_2 X_3 X_4	5	0.74

Exemplo



C_p de Mallow

Este critério avalia o erro quadrático médio dos n valores ajustados segundo um modelo a ser considerado.

Erro de cada valor ajustado é dado por:

$$\hat{Y}_i - \mu_i$$

em que μ_i é o valor verdadeiro da função resposta.

Temos o viés:

$$E(\hat{Y}_i) - \mu_i$$

E um componente aleatório de erro:

$$\hat{Y}_i - E(\hat{Y}_i)$$

C_p de Mallow

$$(\hat{Y}_i - \mu_i)^2 = [(E(\hat{Y}_i) - \mu_i) + (\hat{Y}_i - E(\hat{Y}_i))]^2$$

$$E(\hat{Y}_i - \mu_i)^2 = [E(\hat{Y}_i) - \mu_i]^2 + Var(\hat{Y}_i)$$

Erro quadrático médio total:

$$\sum_{i=1}^{n}[E(\hat{Y}_i) - \mu_i]^2 + \sum_{i=1}^{n} Var(\hat{Y}_i)$$

Medida para o critério:

$$\Gamma_p = \frac{1}{\sigma^2}\left[\sum_{i=1}^{n}[E(\hat{Y}_i) - \mu_i]^2 + \sum_{i=1}^{n} Var(\hat{Y}_i)\right]$$

(erro quadrático médio total dividido pela verdadeira variância do erro)

C_p de Mallow

Estamos considerando incluir $p - 1$ variáveis, mas assuma que o número ideal de variáveis a serem incluídas no modelo seja $P - 1 > p - 1$.

Se assumirmos que o modelo incluindo as $P - 1$ variáveis é correto, temos que $QME(X_1, \ldots, X_{P-1})$ é um estimador não viesado para σ^2.

Estimador para Γ_p é dado por:

$$C_p = \frac{SQE_p}{QME(X_1, \ldots, X_{P-1})} - (n - 2p)$$

C_p de Mallow

Se o modelo com $p-1$ variáveis é adequado, então $E\left[\frac{SQE_p}{(n-p)}\right]=\sigma^2$, de maneira que $E\left[\frac{SQE_p}{QME(X_1,\ldots,X_{P-1})}\right]=n-p$.

Portanto, se o modelo com $p-1$ variáveis é aproximadamente adequado, esperamos que $C_p \approx p$.

Procuramos o menor C_p tal que $C_p \approx p$.

Exemplo

Modelo considerando as variáveis X_1, X_2, X_3 e X_4 ($P - 1 = 4$)

Incluindo apenas X_4 ($p = 2$):

$$C_p = \frac{SQE(X_4)}{QME(X_1, \dots, X_4)} - (n - 2p)$$

Exemplo

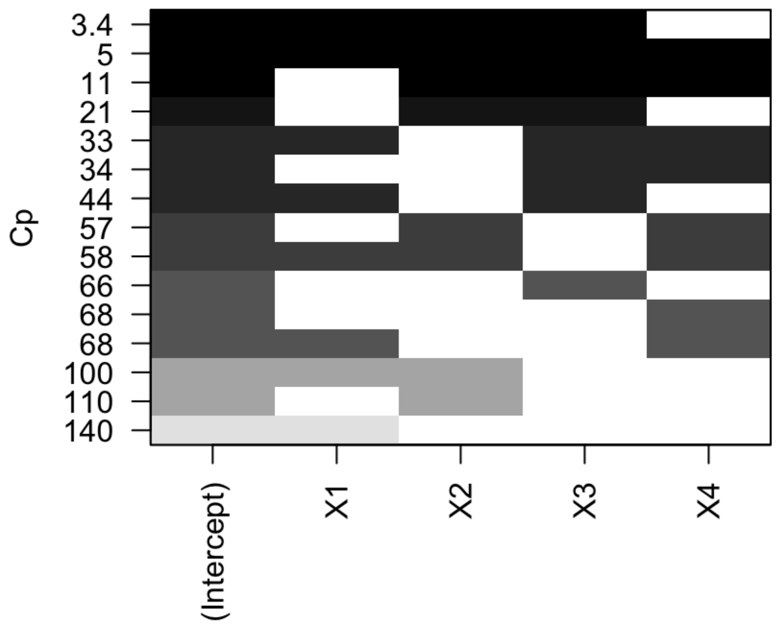
```
## Analysis of Variance Table
##
## Response: lnY
##           Df Sum Sq Mean Sq F value    Pr(>F)
## X4         1 5.3990  5.3990  37.894 1.092e-07 ***
## Residuals 52 7.4087  0.1425
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## Analysis of Variance Table
##
## Response: lnY
##           Df Sum Sq Mean Sq  F value    Pr(>F)
## X1         1 0.7763  0.7763  12.3337 0.0009661 ***
## X2         1 2.5888  2.5888  41.1325 5.377e-08 ***
## X3         1 6.3341  6.3341 100.6408 1.810e-13 ***
## X4         1 0.0246  0.0246   0.3905 0.5349320
## Residuals 49 3.0840  0.0629
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$$C_p = \frac{SQE(X_4)}{QME(X_1, \ldots, X_4)} - (n - 2p) = \frac{7.4087314}{0.062938} - (54 - 2 \times 2) = 67.7147725$$

Exemplo

```
library(leaps)
leaps<-regsubsets(lnY~X1+X2+X3+X4,data=dados,nbest=10)
plot(leaps,scale="Cp")
```



AIC e BIC

Procuramos modelos com valores pequenos de AIC, BIC.

AIC: *Akaike's information criterion*

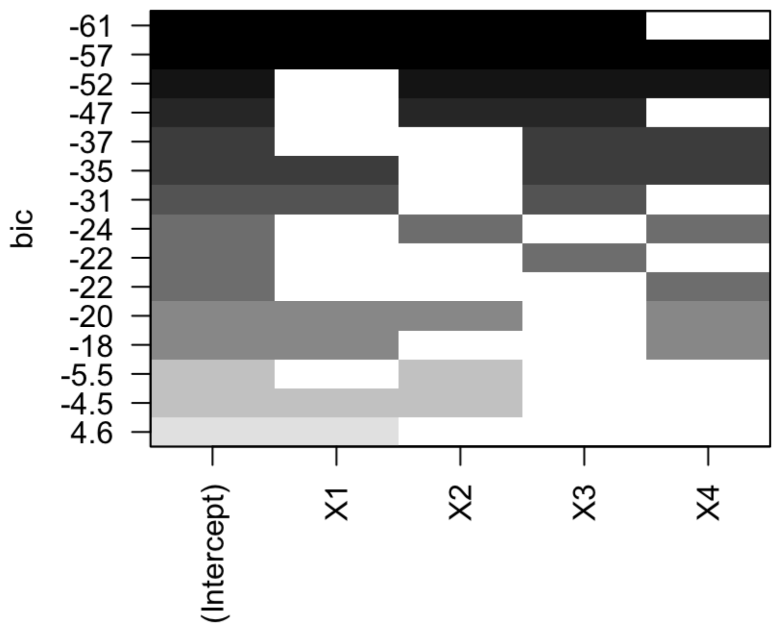
$$AIC_p = n \ln(SQE_p) - n \ln n + 2p$$

BIC: *Bayesian information criterion*

$$BIC_p = n \ln(SQE_p) - n \ln n + \ln(n)p$$

Exemplo

```r
plot(leaps,scale="bic")
```



$PRESS_p$

$PRESS_p$ (*prediction sum of squares*): critério para medir quão adequado é o uso dos valores ajustados obtidos a partir de um modelo com menos variáveis para predizer os valores observados de Y.

$SQE = \sum(Y_i - \hat{Y}_i)^2$ também serve para este propósito.

A diferença é que a medida $PRESS$ é obtida após a exclusão da i-ésima observação e estimação do modelo com as $n-1$ observações restantes, e então usar este modelo para predizer o valor de Y para a i-ésima observação.

Notação: $\hat{Y}_{i(i)}$ indica o valor predito para a i-ésima observação quando esta foi excluída na obtenção do modelo.

$PRESS_p$

$$PRESS_p = \sum_{i=1}^{n} (Y_i - \hat{Y}_{i(i)})^2$$

Modelos com $PRESS_p$ pequenos são considerados bons candidatos (com erro de predição pequeno).

Não é preciso ajustar $n - 1$ vezes o modelo para calcular o $PRESS_p$.

Seja $d_i = Y_i - \hat{Y}_{i(i)}$, reescrevemos: $d_i = \frac{e_i}{1-h_{ii}}$

em que e_i é o resíduo para a i-ésima observação e h_{ii} é o i-ésimo elemento da diagonal de $\mathbf{H} = \mathbf{X}^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}$, obtidos a partir do modelo de regressão com todas as observações incluídas.

Exemplo

```
library(qpcR)
modelo1 <- lm(lnY ~ 1,data=dados)
modelo2 <- lm(lnY ~ X1,data=dados)
modelo3 <- lm(lnY ~ X2,data=dados)
modelo4 <- lm(lnY ~ X3,data=dados)
modelo5 <- lm(lnY ~ X4,data=dados)
modelo6 <- lm(lnY ~ X1+X2,data=dados)
modelo7 <- lm(lnY ~ X1+X3,data=dados)
modelo8 <- lm(lnY ~ X1+X4,data=dados)
modelo9 <- lm(lnY ~ X2+X3,data=dados)
modelo10 <- lm(lnY ~ X2+X4,data=dados)
modelo11 <- lm(lnY ~ X3+X4,data=dados)
modelo12 <- lm(lnY ~ X1+X2+X3,data=dados)
modelo13 <- lm(lnY ~ X1+X2+X4,data=dados)
modelo14 <- lm(lnY ~ X1+X3+X4,data=dados)
modelo15 <- lm(lnY ~ X2+X3+X4,data=dados)
modelo16 <- lm(lnY ~ X1+X2+X3+X4,data=dados)
PRESS(modelo1,verbose=FALSE)$stat
```

```
## [1] 13.2956
```

Exemplo

Variáveis no modelo	p	$PRESS_p$	Variáveis no modelo	p	$PRESS_p$
nenhuma	1	13.296	X_2 X_3	3	5.065
X_1	2	13.512	X_2 X_4	3	7.476
X_2	2	10.744	X_3 X_4	3	6.121
X_3	2	8.327	X_1 X_2 X_3	4	3.914
X_4	2	8.025	X_1 X_2 X_4	4	7.903
X_1 X_2	3	11.062	X_1 X_3 X_4	4	6.207
X_1 X_3	3	6.988	X_2 X_3 X_4	4	4.597
X_1 X_4	3	8.472	X_1 X_2 X_3 X_4	5	4.069

Procedimentos Automáticos para Seleção de Modelos

“Best” Subsets Algorithms

Para o exemplo visto anteriormente, se considerarmos todas as variáveis, temos $2^8 = 256$ modelos possíveis.

Exemplo - Usando AIC_p

```r
library(bestglm)
Xy = dados[,-9] # excluindo coluna do Y original, usamos ln(Y) como variável resposta
names(Xy) <- c(names(Xy)[1:8],"y")
modelos <- bestglm(Xy,IC="AIC",TopModels = 2)
modelos$Subsets
```

```
##    (Intercept)    X1    X2    X3    X4    X5    X6    X7    X8 logLikelihood
## 0         TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE      38.85126
## 1         TRUE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE      53.91343
## 2         TRUE FALSE  TRUE  TRUE FALSE FALSE FALSE FALSE FALSE      68.24165
## 3         TRUE FALSE  TRUE  TRUE FALSE FALSE FALSE FALSE  TRUE      79.49246
## 4         TRUE  TRUE  TRUE  TRUE FALSE FALSE FALSE FALSE  TRUE      86.67568
## 5         TRUE  TRUE  TRUE  TRUE FALSE FALSE  TRUE FALSE  TRUE      87.90259
## 6*        TRUE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE FALSE  TRUE      88.91714
## 7         TRUE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE      89.36782
## 8         TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE      89.38549
##            AIC
## 0    -77.70252
## 1   -105.82686
## 2   -132.48329
## 3   -152.98493
## 4   -165.35135
## 5   -165.80517
## 6*  -165.83429
## 7   -164.73565
## 8   -162.77098
```

Exemplo - Usando AIC_p

```
melhor <- which(modelos$Subsets$AIC==min(modelos$Subsets$AIC))
numvar <- dim(Xy)[2]-1 # total de variáveis consideradas inicialmente
varincluidas <- modelos$Subsets[melhor,2:(numvar+1)] # variaveis escolhidas segundo criterio
varincluidas
```

```
##      X1   X2   X3    X4   X5   X6    X7   X8
## 6* TRUE TRUE TRUE FALSE TRUE TRUE FALSE TRUE
```

```
modeloescolhidoAIC <- lm(y ~ .,data=Xy[,c(which(varincluidas==TRUE),which(names(Xy)=="y"))])
summary(modeloescolhidoAIC)$coef
```

```
##                 Estimate  Std. Error   t value     Pr(>|t|)
## (Intercept)  4.053974209 0.234793506 17.266126 5.572016e-22
## X1           0.071517057 0.018637294  3.837309 3.701898e-04
## X2           0.013755482 0.001709437  8.046792 2.169036e-10
## X3           0.015116499 0.001385313 10.911972 1.777375e-14
## X5          -0.003450094 0.002571776 -1.341522 1.861972e-01
## X6           0.087316639 0.057701672  1.513243 1.369140e-01
## X8           0.350903932 0.076391406  4.593500 3.276184e-05
```

Exemplo - Usando BIC_p

```r
modelos <- bestglm(Xy,IC="BIC")
modelos$Subsets
```

```
##    (Intercept)    X1    X2    X3    X4    X5    X6    X7    X8 logLikelihood
## 0         TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE      38.85126
## 1         TRUE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE      53.91343
## 2         TRUE FALSE  TRUE  TRUE FALSE FALSE FALSE FALSE FALSE      68.24165
## 3         TRUE FALSE  TRUE  TRUE FALSE FALSE FALSE FALSE  TRUE      79.49246
## 4*        TRUE  TRUE  TRUE  TRUE FALSE FALSE FALSE FALSE  TRUE      86.67568
## 5         TRUE  TRUE  TRUE  TRUE FALSE FALSE  TRUE FALSE  TRUE      87.90259
## 6         TRUE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE FALSE  TRUE      88.91714
## 7         TRUE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE      89.36782
## 8         TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE      89.38549
##             BIC
## 0     -77.70252
## 1    -103.83788
## 2    -128.50532
## 3    -147.01798
## 4*   -157.39542
## 5    -155.86025
## 6    -153.90039
## 7    -150.81276
## 8    -146.85911
```

Exemplo - Usando BIC_p

```
melhor <- which(modelos$Subsets$BIC==min(modelos$Subsets$BIC))
varincluidas <- modelos$Subsets[melhor,2:(numvar+1)] # variaveis escolhidas segundo criterio
varincluidas
```

```
##       X1   X2   X3    X4    X5    X6    X7   X8
## 4* TRUE TRUE TRUE FALSE FALSE FALSE FALSE TRUE
```

```
modeloescolhidoBIC <- lm(y ~ .,data=Xy[,c(which(varincluidas==TRUE),which(names(Xy)=="y"))])
summary(modeloescolhidoBIC)$coef
```

```
##               Estimate  Std. Error   t value     Pr(>|t|)
## (Intercept) 3.85241856 0.192695224 19.992289 3.279284e-25
## X1          0.07332263 0.018973044  3.864569 3.273887e-04
## X2          0.01418507 0.001730632  8.196469 9.581863e-11
## X3          0.01545270 0.001395609 11.072371 6.145977e-15
## X8          0.35296762 0.077190626  4.572675 3.290701e-05
```

Exemplo - Usando $PRESS_p$

```
modelos <- bestglm(Xy,IC="LOOCV")
modelos$Subsets
```

```
##    (Intercept)    X1    X2    X3    X4    X5    X6    X7    X8 logLikelihood
## 0         TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE      38.85126
## 1         TRUE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE      53.91343
## 2         TRUE FALSE  TRUE  TRUE FALSE FALSE FALSE FALSE FALSE      68.24165
## 3         TRUE FALSE  TRUE  TRUE FALSE FALSE FALSE FALSE  TRUE      79.49246
## 4*        TRUE  TRUE  TRUE  TRUE FALSE FALSE FALSE FALSE  TRUE      86.67568
## 5         TRUE  TRUE  TRUE  TRUE FALSE FALSE  TRUE FALSE  TRUE      87.90259
## 6         TRUE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE FALSE  TRUE      88.91714
## 7         TRUE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE      89.36782
## 8         TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE      89.38549
##         LOOCV
## 0  0.24621473
## 1  0.15419845
## 2  0.09380257
## 3  0.06424821
## 4* 0.05069947
## 5  0.05153172
## 6  0.05133936
## 7  0.05201306
## 8  0.05428207
```

Exemplo - Usando $PRESS_p$

```
melhor <- which(modelos$Subsets$LOOCV==min(modelos$Subsets$LOOCV))
varincluidas <- modelos$Subsets[melhor,2:(numvar+1)] # variaveis escolhidas segundo criterio
varincluidas
```

```
##       X1   X2   X3    X4    X5    X6    X7   X8
## 4* TRUE TRUE TRUE FALSE FALSE FALSE FALSE TRUE
```

```
modeloescolhidoPRESS <- lm(y ~ .,data=Xy[,c(which(varincluidas==TRUE),which(names(Xy)=="y"))])
summary(modeloescolhidoPRESS)$coef
```

```
##               Estimate  Std. Error   t value     Pr(>|t|)
## (Intercept) 3.85241856 0.192695224 19.992289 3.279284e-25
## X1          0.07332263 0.018973044  3.864569 3.273887e-04
## X2          0.01418507 0.001730632  8.196469 9.581863e-11
## X3          0.01545270 0.001395609 11.072371 6.145977e-15
## X8          0.35296762 0.077190626  4.572675 3.290701e-05
```

Exemplo - C_p de Mallow, R^2_p, $R^2_{a,p}$ e BIC_p

```
library(leaps)
modelos <- regsubsets(y ~ .,data=Xy,nbest=2)
resultados = data.frame(cbind("p"=rowSums(summary(modelos)$which),summary(modelos)$which,
                              "Cp"=round(summary(modelos)$cp,2),
                              "R2"=round(summary(modelos)$rsq,2),
                          "R2adj"=round(summary(modelos)$adjr2,2),"BIC"=round(summary(modelos)$bic,2)))
resultados
```

```
##       p X.Intercept. X1 X2 X3 X4 X5 X6 X7 X8     Cp   R2 R2adj    BIC
## X1    2            1  0  0  1  0  0  0  0  0 117.41 0.43  0.42 -22.15
## X1.1  2            1  0  0  0  1  0  0  0  0 119.17 0.42  0.41 -21.58
## X2    3            1  0  1  1  0  0  0  0  0  50.47 0.66  0.65 -46.81
## X2.1  3            1  0  0  1  1  0  0  0  0  69.13 0.60  0.58 -37.44
## X3    4            1  0  1  1  0  0  0  0  1  18.91 0.78  0.76 -65.33
## X3.1  4            1  1  1  1  0  0  0  0  0  24.98 0.76  0.74 -60.50
## X4    5            1  1  1  1  0  0  0  0  1   5.75 0.83  0.82 -75.70
## X4.1  5            1  0  1  1  1  0  0  0  1  10.27 0.81  0.80 -71.01
## X5    6            1  1  1  1  0  0  1  0  1   5.54 0.84  0.82 -74.17
## X5.1  6            1  1  1  1  0  1  0  0  1   6.02 0.84  0.82 -73.63
## X6    7            1  1  1  1  0  1  1  0  1   5.79 0.84  0.82 -72.21
## X6.1  7            1  1  1  1  0  0  1  1  1   7.03 0.84  0.82 -70.76
## X7    8            1  1  1  1  0  1  1  1  1   7.03 0.85  0.82 -69.12
## X7.1  8            1  1  1  1  1  1  1  0  1   7.74 0.84  0.82 -68.28
## X8    9            1  1  1  1  1  1  1  1  1   9.00 0.85  0.82 -65.17
```

Método *Stepwise*

- Método menos intensivo computacionalmente.
- Ao final, obtém-se apenas 1 modelo candidato.
- *forward*, *backward*, *both*

Forward Selection

Início considerando $P-1$ variáveis.

1- Ajuste uma regressão linear simples com cada uma das $P-1$ variáveis. Para cada regressão, calcule a estatística t^* para testar se o coeficiente angular é 0.

$$t_k^* = \frac{\hat{\beta}_k}{\sqrt{\widehat{Var}(\hat{\beta}_k)}}$$

2- Considere a variável cujo $|t^*|$ é o maior. Inclua esta variável caso $|t^*|$ esteja acima de algum valor pré-determinado.

3 - Se alguma variável é incluída, por exemplo, X_7 ajustam-se regressões com pares de variáveis, sendo que sempre uma delas é X_7. Calcula-se t^* para a nova variável incluída e repita o passo 2 para decidir qual a segunda variável a ser incluída no modelo.

4 - Repita até considerar todas as variáveis.

Backward Selection

1. Ajuste uma regressão linear múltipla com todas as $P-1$ variáveis.
2. Teste iterativamente se uma das variáveis pode ser eliminada.

Exemplo: *Forward Regression*

```
completo = lm(y~.,data=Xy)
vazio = lm(y~1, data=Xy)
step(vazio, scope=list(upper=completo, lower=vazio), direction='forward', trace=TRUE)
Start:  AIC=-75.7
y ~ 1

       Df Sum of Sq     RSS      AIC
+ X3    1    5.4762  7.3316 -103.827
+ X4    1    5.3990  7.4087 -103.262
+ X2    1    2.8285  9.9792  -87.178
+ X8    1    1.7798 11.0279  -81.782
+ X1    1    0.7763 12.0315  -77.079
+ X6    1    0.6897 12.1180  -76.692
<none>              12.8077  -75.703
+ X5    1    0.2691 12.5386  -74.849
+ X7    1    0.2052 12.6025  -74.575
```

Exemplo: *Forward Regression*

```
Step:  AIC=-103.83
y ~ X3

       Df Sum of Sq    RSS     AIC
+ X2    1   3.01908 4.3125 -130.48
+ X4    1   2.20187 5.1297 -121.11
+ X1    1   1.55061 5.7810 -114.66
+ X8    1   1.13756 6.1940 -110.93
<none>              7.3316 -103.83
+ X6    1   0.25854 7.0730 -103.77
+ X5    1   0.23877 7.0928 -103.61
+ X7    1   0.06498 7.2666 -102.31
```

Exemplo: *Forward Regression*

```
Step:  AIC=-130.48
y ~ X3 + X2

       Df Sum of Sq    RSS     AIC
+ X8    1   1.46961 2.8429 -150.99
+ X1    1   1.20395 3.1085 -146.16
+ X4    1   0.69836 3.6141 -138.02
+ X7    1   0.22632 4.0862 -131.39
+ X5    1   0.16461 4.1479 -130.59
<none>              4.3125 -130.48
+ X6    1   0.08245 4.2300 -129.53
```

Exemplo: *Forward Regression*

```
Step:  AIC=-150.98
y ~ X3 + X2 + X8

       Df Sum of Sq    RSS     AIC
+ X1    1   0.66408 2.1788 -163.35
+ X4    1   0.46630 2.3766 -158.66
+ X6    1   0.13741 2.7055 -151.66
<none>              2.8429 -150.99
+ X5    1   0.07081 2.7721 -150.35
+ X7    1   0.02464 2.8182 -149.46
```

Exemplo: *Forward Regression*

```
 Step:  AIC=-163.35
y ~ X3 + X2 + X8 + X1

       Df Sum of Sq    RSS     AIC
+ X6    1  0.096791 2.0820 -163.81
<none>                2.1788 -163.35
+ X5    1  0.075876 2.1029 -163.26
+ X4    1  0.041701 2.1371 -162.40
+ X7    1  0.022944 2.1559 -161.92
```

Exemplo: *Forward Regression*

```
Step:  AIC=-163.81
y ~ X3 + X2 + X8 + X1 + X6

       Df Sum of Sq    RSS     AIC
+ X5    1  0.076782 2.0052 -163.83
<none>              2.0820 -163.81
+ X7    1  0.022387 2.0596 -162.39
+ X4    1  0.016399 2.0656 -162.23
```

Exemplo: *Forward Regression*

```
Step:  AIC=-163.83
y ~ X3 + X2 + X8 + X1 + X6 + X5

       Df Sum of Sq    RSS     AIC
<none>               2.0052 -163.83
+ X7    1  0.033193 1.9720 -162.74
+ X4    1  0.002284 2.0029 -161.90

Call:
lm(formula = y ~ X3 + X2 + X8 + X1 + X6 + X5, data = Xy)

Coefficients:
(Intercept)           X3           X2           X8           X1           X6           X5
    4.05397      0.01512      0.01376      0.35090      0.07152      0.08732     -0.00345
```

Exemplo: *Backward Regression*

```
completo = lm(y~.,data=Xy)
vazio = lm(y~1, data=Xy)
step(completo, scope=list(upper=completo, lower=vazio), direction='backward', trace=TRUE)

Start:  AIC=-160.77
y ~ X1 + X2 + X3 + X4 + X5 + X6 + X7 + X8

       Df Sum of Sq    RSS     AIC
- X4    1   0.00129 1.9720 -162.74
- X7    1   0.03220 2.0029 -161.90
- X5    1   0.07354 2.0443 -160.79
<none>              1.9707 -160.77
- X6    1   0.08415 2.0549 -160.51
- X1    1   0.31809 2.2888 -154.69
- X8    1   0.84573 2.8165 -143.49
- X2    1   2.09045 4.0612 -123.72
- X3    1   2.99085 4.9616 -112.91
```

Exemplo: *Backward Regression*

```
Step:  AIC=-162.74
y ~ X1 + X2 + X3 + X5 + X6 + X7 + X8

       Df Sum of Sq    RSS      AIC
- X7    1    0.0332 2.0052 -163.834
<none>              1.9720 -162.736
- X5    1    0.0876 2.0596 -162.389
- X6    1    0.0971 2.0691 -162.141
- X1    1    0.6267 2.5988 -149.833
- X8    1    0.8446 2.8166 -145.486
- X2    1    2.6731 4.6451 -118.471
- X3    1    5.0986 7.0706  -95.784
```

Exemplo: *Backward Regression*

```
Step:  AIC=-163.83
y ~ X1 + X2 + X3 + X5 + X6 + X8

       Df Sum of Sq    RSS      AIC
<none>              2.0052 -163.834
- X5    1    0.0768 2.0820 -163.805
- X6    1    0.0977 2.1029 -163.265
- X1    1    0.6282 2.6335 -151.117
- X8    1    0.9002 2.9055 -145.809
- X2    1    2.7626 4.7678 -119.064
- X3    1    5.0801 7.0853  -97.672
```

Exemplo: *Backward Regression*

```
Call:
lm(formula = y ~ X1 + X2 + X3 + X5 + X6 + X8, data = Xy)

Coefficients:
(Intercept)           X1           X2           X3           X5           X6           X8
    4.05397      0.07152      0.01376      0.01512     -0.00345      0.08732      0.35090
```

Exemplo: *Forward Stepwise Regression*

```
completo = lm(y~.,data=Xy)
vazio = lm(y~1, data=Xy)
step(vazio, scope=list(upper=completo, lower=vazio), direction='both', trace=TRUE)

Start:  AIC=-75.7
y ~ 1

       Df Sum of Sq     RSS      AIC
+ X3    1    5.4762  7.3316 -103.827
+ X4    1    5.3990  7.4087 -103.262
+ X2    1    2.8285  9.9792  -87.178
+ X8    1    1.7798 11.0279  -81.782
+ X1    1    0.7763 12.0315  -77.079
+ X6    1    0.6897 12.1180  -76.692
<none>               12.8077  -75.703
+ X5    1    0.2691 12.5386  -74.849
+ X7    1    0.2052 12.6025  -74.575
```

Exemplo: *Forward Stepwise Regression*

```
Step:  AIC=-103.83
y ~ X3

       Df Sum of Sq     RSS      AIC
+ X2    1    3.0191  4.3125 -130.483
+ X4    1    2.2019  5.1297 -121.113
+ X1    1    1.5506  5.7810 -114.658
+ X8    1    1.1376  6.1940 -110.932
<none>                7.3316 -103.827
+ X6    1    0.2585  7.0730 -103.765
+ X5    1    0.2388  7.0928 -103.615
+ X7    1    0.0650  7.2666 -102.308
- X3    1    5.4762 12.8077  -75.703
```

Exemplo: *Forward Stepwise Regression*

```
Step:  AIC=-130.48
y ~ X3 + X2

       Df Sum of Sq    RSS      AIC
+ X8    1    1.4696 2.8429 -150.985
+ X1    1    1.2040 3.1085 -146.161
+ X4    1    0.6984 3.6141 -138.023
+ X7    1    0.2263 4.0862 -131.394
+ X5    1    0.1646 4.1479 -130.585
<none>              4.3125 -130.483
+ X6    1    0.0824 4.2300 -129.526
- X2    1    3.0191 7.3316 -103.827
- X3    1    5.6667 9.9792  -87.178
```

Exemplo: *Forward Stepwise Regression*

```
Step:  AIC=-150.98
y ~ X3 + X2 + X8

       Df Sum of Sq    RSS      AIC
+ X1    1    0.6641 2.1788 -163.351
+ X4    1    0.4663 2.3766 -158.659
+ X6    1    0.1374 2.7055 -151.660
<none>              2.8429 -150.985
+ X5    1    0.0708 2.7721 -150.347
+ X7    1    0.0246 2.8182 -149.455
- X8    1    1.4696 4.3125 -130.483
- X2    1    3.3511 6.1940 -110.932
- X3    1    4.9456 7.7885  -98.562
```

Exemplo: *Forward Stepwise Regression*

```
Step:  AIC=-163.35
y ~ X3 + X2 + X8 + X1

       Df Sum of Sq    RSS      AIC
+ X6    1    0.0968 2.0820 -163.805
<none>              2.1788 -163.351
+ X5    1    0.0759 2.1029 -163.265
+ X4    1    0.0417 2.1371 -162.395
+ X7    1    0.0229 2.1559 -161.923
- X1    1    0.6641 2.8429 -150.985
- X8    1    0.9297 3.1085 -146.161
- X2    1    2.9873 5.1661 -118.731
- X3    1    5.4513 7.6301  -97.671
```

Exemplo: *Forward Stepwise Regression*

```
Step:  AIC=-163.81
y ~ X3 + X2 + X8 + X1 + X6

       Df Sum of Sq    RSS      AIC
+ X5    1    0.0768 2.0052 -163.834
<none>              2.0820 -163.805
- X6    1    0.0968 2.1788 -163.351
+ X7    1    0.0224 2.0596 -162.389
+ X4    1    0.0164 2.0656 -162.232
- X1    1    0.6235 2.7055 -151.660
- X8    1    0.9745 3.0565 -145.072
- X2    1    2.8268 4.9088 -119.490
- X3    1    5.0791 7.1611  -99.097
```

Exemplo: *Forward Stepwise Regression*

```
Step:  AIC=-163.83
y ~ X3 + X2 + X8 + X1 + X6 + X5

       Df Sum of Sq    RSS      AIC
<none>              2.0052 -163.834
- X5    1    0.0768 2.0820 -163.805
- X6    1    0.0977 2.1029 -163.265
+ X7    1    0.0332 1.9720 -162.736
+ X4    1    0.0023 2.0029 -161.896
- X1    1    0.6282 2.6335 -151.117
- X8    1    0.9002 2.9055 -145.809
- X2    1    2.7626 4.7678 -119.064
- X3    1    5.0801 7.0853  -97.672

Call:
lm(formula = y ~ X3 + X2 + X8 + X1 + X6 + X5, data = Xy)

Coefficients:
(Intercept)           X3           X2           X8           X1           X6           X5
    4.05397      0.01512      0.01376      0.35090      0.07152      0.08732     -0.00345
```

Validação de Modelos

Introdução

Verificar se um modelo candidato tem bom desempenho em dados independentes daqueles usados para ajuste.

- Coletar novos dados para verificar o modelo e seu poder preditivo.
- Deixar parte dos dados de fora do ajuste, para usar na validação.

Validação Cruzada

Quando temos um grande número de observações, podemos dividir os dados em duas partes: *treinamento* e *teste*.

Com o subconjunto *treinamento* ajustamos o modelo.

Com o subconjunto *teste* verificamos o poder preditivo do modelo.

Calculamos o *mean squared prediction error*:

$$MSPR = \frac{\sum_{i=1}^{n^*}(Y_i - \hat{Y}_i)^2}{n^*}$$

em que Y_i é o valor da variável resposta da i-ésima observação do conjunto teste, $\hat{Y}_i$ é o valor predito para a i-ésima observação do conjunto teste segundo o modelo usando o conjunto treinamento e n^* é o total de observações no conjunto teste.

Exemplo

Temos 54 observações que não foram utilizadas na escolha do modelo para os dados sobre cirurgia. Este será o conjunto de dados *teste*.

Com os dados de *treinamento*, obtemos, usando $PRESS_p$ e BIC_p:

Modelo 1:

$$\ln(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_8 X_8 + \varepsilon$$

Usando C_p, temos o Modelo 2:

$$\ln(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_5 X_5 + \beta_8 X_8 + \varepsilon$$

Usando AIC_p e $R^2_{a,p}$, temos o Modelo 3:

$$\ln(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_5 X_5 + \beta_6 X_6 + \beta_8 X_8 + \varepsilon$$

Exemplo - Modelo 1

```r
dadosT <- read.table("./dados/CH09TA05.txt")
colnames(dadosT) <- c("X1","X2","X3","X4","X5","X6","X7","X8","Y","lnY")
modelo1 <- lm(lnY ~ X1 + X2 + X3 + X8,data=dados)
yhat <- predict(modelo1,newdata=dadosT)
MSPR <- function(yhat,yobs){
  mean((yobs-yhat)^2)
}
```

Variável	Estimativa	Erro-Padrão
Intercepto	3.8524186	0.1926952
X_1	0.0733226	0.018973
X_2	0.0141851	0.0017306
X_3	0.0154527	0.0013956
X_8	0.3529676	0.0771906

MSE é 0.044 e $MSPR$ é 0.077

Exemplo - Modelo 2

```
modelo2 <- lm(lnY ~ X1 + X2 + X3 + X5 + X8,data=dados)
yhat <- predict(modelo2,newdata=dadosT)
```

Variável	Estimativa	Erro-Padrão
Intercepto	4.0381206	0.2376904
X_1	0.0736065	0.0188341
X_2	0.0140523	0.0017208
X_3	0.0154557	0.0013853
X_5	-0.0034296	0.0026061
X_8	0.3412188	0.0771389

MSE é 0.044 e $MSPR$ é 0.08

Exemplo - Modelo 3

```
modelo3 <- lm(lnY ~ X1 + X2 + X3 + X5 + X6 + X8,data=dados)
yhat <- predict(modelo3,newdata=dadosT)
```

Variável	Estimativa	Erro-Padrão
Intercepto	4.0539742	0.2347935
X_1	0.0715171	0.0186373
X_2	0.0137555	0.0017094
X_3	0.0151165	0.0013853
X_5	-0.0034501	0.0025718
X_6	0.0873166	0.0577017
X_8	0.3509039	0.0763914

MSE é 0.043 e $MSPR$ é 0.079

Agradecimento

- Slides criados por Samara F Kiihl / IMECC / UNICAMP

Leitura

- Applied Linear Statistical Models: Capítulo 9.
- Faraway - Linear Models with R: Capítulo 10
- Draper & Smith - Applied Regression Analysis: Capítulo 15.
- Tutorial: Model Selection in R
- `bestglm`