AULA 8: SELEÇÕES DE MODELOS / EXTENSÕES DOS MODELOS LINEARES

Análise Estatística e Modelagem de Dados Ecológicos

Thiago S. F. Silva - tsfsilva@rc.unesp.br

1 de Abril de 2015

Programa de Pós Graduação em Ecologia e Biodiversidade - UNESP

OUTLINE

SELEÇÃO DE VARIÁVEIS E CONSTRUÇÃO DO MODELO

Atualmente, muito da pesquisa científica e análise de dados se baseia no modelo simplista de testes contra uma hipótese nula ”ingênua”(*naive*) e pouco informativa, e que resulta de uma amálgama entre os métodos propostos por Fisher e Neyman-Pearson.

Vários autores sugerem uma estratégia diferente, onde múltiplas hipóteses podem ser avaliadas comparativamente, a partir da análise da evidencia a favor de cada hipótese.

Cada hipótese pode ser expressa em termos de um modelo estatístico, e a comparação entre modelos oferecerá a evidência que suporta ou não as diferentes hipóteses

Esse processo pode ser chamado de **Construção do Modelo**.

O PROCESSO DE CONSTRUÇÃO DO MODELO

1) Construa o seu modelo conceitual

- O que já se sabe sobre o seu problema (arcabouço teórico)?

1) Construa o seu modelo conceitual

- O que já se sabe sobre o seu problema (arcabouço teórico)?
- Com base nesse conhecimento, que tipo de relações você espera?

1) Construa o seu modelo conceitual

- O que já se sabe sobre o seu problema (arcabouço teórico)?
- Com base nesse conhecimento, que tipo de relações você espera?
- Essas relações já foram quantificadas em outros estudos, mesmo que para outros sistemas?

1) Construa o seu modelo conceitual

- O que já se sabe sobre o seu problema (arcabouço teórico)?
- Com base nesse conhecimento, que tipo de relações você espera?
- Essas relações já foram quantificadas em outros estudos, mesmo que para outros sistemas?
- O método "escopeta"(*shotgun*) é, na maioria das vezes, uma perda de tempo e recursos

1) Construa o seu modelo conceitual

- O que já se sabe sobre o seu problema (arcabouço teórico)?
- Com base nesse conhecimento, que tipo de relações você espera?
- Essas relações já foram quantificadas em outros estudos, mesmo que para outros sistemas?
- O método ”escopeta”(*shotgun*) é, na maioria das vezes, uma perda de tempo e recursos
- Você pode gastar o mesmo dinheiro/esforço para coletar 10 réplicas de 10 variáveis...ou 100 réplicas de uma variável

1) Construa o seu modelo conceitual

- O que já se sabe sobre o seu problema (arcabouço teórico)?
- Com base nesse conhecimento, que tipo de relações você espera?
- Essas relações já foram quantificadas em outros estudos, mesmo que para outros sistemas?
- O método "escopeta"(*shotgun*) é, na maioria das vezes, uma perda de tempo e recursos
- Você pode gastar o mesmo dinheiro/esforço para coletar 10 réplicas de 10 variáveis...ou 100 réplicas de uma variável
- E não esqueça do esforço de análise!

O PROCESSO DE CONSTRUÇÃO DO MODELO

1) Construa o seu modelo conceitual.

2)

O PROCESSO DE CONSTRUÇÃO DO MODELO

1) Construa o seu modelo conceitual.

2) Traduza suas hipóteses em modelos estatísticos.

- Quais são as hipóteses?

O PROCESSO DE CONSTRUÇÃO DO MODELO

1) Construa o seu modelo conceitual.
2) Traduza suas hipóteses em modelos estatísticos.

- Quais são as hipóteses?
- Que informações (variáveis) preciso medir para testar essas hipóteses?

O PROCESSO DE CONSTRUÇÃO DO MODELO

1) Construa o seu modelo conceitual.
2) Traduza suas hipóteses em modelos estatísticos.

- Quais são as hipóteses?

- Que informações (variáveis) preciso medir para testar essas hipóteses?

- O desenho experimental é essencial!

O PROCESSO DE CONSTRUÇÃO DO MODELO

1) Construa o seu modelo conceitual

2) Traduza suas hipóteses em modelos estatísticos

- Experimentos x estudos observacionais (confirmatórios ou exploratórios).

O PROCESSO DE CONSTRUÇÃO DO MODELO

1) Construa o seu modelo conceitual
2) Traduza suas hipóteses em modelos estatísticos

- Experimentos x estudos observacionais (confirmatórios ou exploratórios).
- O seu modelo só é valido dentro do escopo das variáveis preditivas.

O PROCESSO DE CONSTRUÇÃO DO MODELO

1) Construa o seu modelo conceitual
2) Traduza suas hipóteses em modelos estatísticos

- Experimentos x estudos observacionais (confirmatórios ou exploratórios).
- O seu modelo só é valido dentro do escopo das variáveis preditivas.
- Generalidade x especificidade.

O PROCESSO DE CONSTRUÇÃO DO MODELO

1) Construa o seu modelo conceitual
2) Traduza suas hipóteses em modelos estatísticos

- Experimentos x estudos observacionais (confirmatórios ou exploratórios).
- O seu modelo só é valido dentro do escopo das variáveis preditivas.
- Generalidade x especificidade.
- Tamanho da amostra: qual o tamanho do efeito que você deseja detectar, e com qual nivel de confiança?

1) Construa o seu modelo conceitual
2) Traduza suas hipóteses em modelos estatísticos

- Experimentos x estudos observacionais (confirmatórios ou exploratórios).
- O seu modelo só é valido dentro do escopo das variáveis preditivas.
- Generalidade x especificidade.
- Tamanho da amostra: qual o tamanho do efeito que você deseja detectar, e com qual nivel de confiança?
- Estudos pilotos e ou simulações podem fazer toda a diferença.

O PROCESSO DE CONSTRUÇÃO DO MODELO

1) Construa o seu modelo conceitual

2) Traduza suas hipóteses em modelos estatísticos

3) Colete seus dados

O PROCESSO DE CONSTRUÇÃO DO MODELO

1) Construa o seu modelo conceitual
2) Traduza suas hipóteses em modelos estatísticos
3) Colete seus dados
4) Faça uma análise exploratória

1) Construa o seu modelo conceitual

2) Traduza suas hipóteses em modelos estatísticos

3) Colete seus dados

4) Faça uma análise exploratória

- Estatísticas descritivas (médias, desvios, quantis)

O PROCESSO DE CONSTRUÇÃO DO MODELO

1) Construa o seu modelo conceitual

2) Traduza suas hipóteses em modelos estatísticos

3) Colete seus dados

4) Faça uma análise exploratória

- Estatísticas descritivas (médias, desvios, quantis)
- Histogramas, boxplots, curvas de densidade de distribuição

1) Construa o seu modelo conceitual

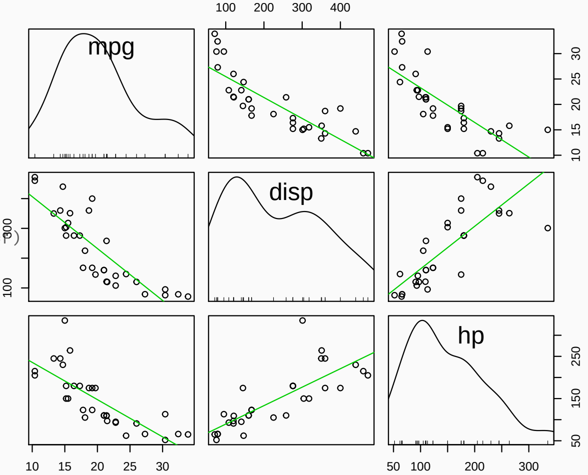
2) Traduza suas hipóteses em modelos estatísticos

3) Colete seus dados

4) Faça uma análise exploratória

- Estatísticas descritivas (médias, desvios, quantis)
- Histogramas, boxplots, curvas de densidade de distribuição
- Gráficos de dispersão (*scatterplots*) e correlações:
  - Matrizes de gráficos de dispersão
  - Matrizes de correlação
- Este passo deve ser uma pré-analise dos dados frente às hipóteses pré-estabelecidas, e não o início do processo de formulação de hipóteses!

```
data(mtcars)
print(cor(mtcars),digits=2)

##        mpg   cyl  disp    hp   drat    wt   qsec    vs     am  gear   carb
## mpg   1.00 -0.85 -0.85 -0.78  0.681 -0.87  0.419  0.66  0.600  0.48 -0.551
## cyl  -0.85  1.00  0.90  0.83 -0.700  0.78 -0.591 -0.81 -0.523 -0.49  0.527
## disp -0.85  0.90  1.00  0.79 -0.710  0.89 -0.434 -0.71 -0.591 -0.56  0.395
## hp   -0.78  0.83  0.79  1.00 -0.449  0.66 -0.708 -0.72 -0.243 -0.13  0.750
## drat  0.68 -0.70 -0.71 -0.45  1.000 -0.71  0.091  0.44  0.713  0.70 -0.091
## wt   -0.87  0.78  0.89  0.66 -0.712  1.00 -0.175 -0.55 -0.692 -0.58  0.428
## qsec  0.42 -0.59 -0.43 -0.71  0.091 -0.17  1.000  0.74 -0.230 -0.21 -0.656
## vs    0.66 -0.81 -0.71 -0.72  0.440 -0.55  0.745  1.00  0.168  0.21 -0.570
## am    0.60 -0.52 -0.59 -0.24  0.713 -0.69 -0.230  0.17  1.000  0.79  0.058
## gear  0.48 -0.49 -0.56 -0.13  0.700 -0.58 -0.213  0.21  0.794  1.00  0.274
## carb -0.55  0.53  0.39  0.75 -0.091  0.43 -0.656 -0.57  0.058  0.27  1.000
```




```r
library(car)
scatterplotMatrix(mtcars[,c(1,3:4)],smoother=F)
 smoother = F desativa a opção de linha suavizada
```

1) Construa o seu modelo conceitual

2) Traduza suas hipóteses em modelos estatísticos

3) Colete seus dados

4) Faça uma análise exploratória

5) “Seleção de variáveis” - Avaliação das hipóteses

O PROCESSO DE CONSTRUÇÃO DO MODELO

1) Construa o seu modelo conceitual

2) Traduza suas hipóteses em modelos estatísticos

3) Colete seus dados

4) Faça uma análise exploratória

5) “Seleção de variáveis” - Avaliação das hipóteses

- Aqui, os seus modelos serão avaliados comparativamente

1) Construa o seu modelo conceitual

2) Traduza suas hipóteses em modelos estatísticos

3) Colete seus dados

4) Faça uma análise exploratória

5) "Seleção de variáveis" - Avaliação das hipóteses

- Aqui, os seus modelos serão avaliados comparativamente
- Na maioria das vezes, as hipóteses são formuladas *post hoc*, e as variáveis são adicionadas e removidas durante esse processo. Não é o ideal.

O PROCESSO DE CONSTRUÇÃO DO MODELO

1) Construa o seu modelo conceitual

2) Traduza suas hipóteses em modelos estatísticos

3) Colete seus dados

4) Faça uma análise exploratória

5) “Seleção de variáveis” - Avaliação das hipóteses

- Aqui, os seus modelos serão avaliados comparativamente
- Na maioria das vezes, as hipóteses são formuladas *post hoc*, e as variáveis são adicionadas e removidas durante esse processo. Não é o ideal.
- Você pretende **explicar** a relação entre as variáveis? Comece com o modelo completo e avalie os coeficientes e incertezas (erros e valores p).

O PROCESSO DE CONSTRUÇÃO DO MODELO

1) Construa o seu modelo conceitual
2) Traduza suas hipóteses em modelos estatísticos
3) Colete seus dados
4) Faça uma análise exploratória
5) “Seleção de variáveis” - Avaliação das hipóteses

- Cuidado com o efeito da multicolinearidade! Calcule os VIFs, e aplique medidas corretivas se necessário!

O PROCESSO DE CONSTRUÇÃO DO MODELO

1) Construa o seu modelo conceitual

2) Traduza suas hipóteses em modelos estatísticos

3) Colete seus dados

4) Faça uma análise exploratória

5) “Seleção de variáveis” - Avaliação das hipóteses

- Cuidado com o efeito da multicolinearidade! Calcule os VIFs, e aplique medidas corretivas se necessário!
- O objetivo é prever? Avalie a contribuição de cada variável para o modelo final, e mantenha só as mais importantes. Esta avaliação é mais robusta se houver validação independente.

Para cada conjunto de $k = p - 1$ preditores, existem 2^{p-1} combinações de variáveis.

Para cada conjunto de $k = p - 1$ preditores, existem 2^{p-1} combinações de variáveis.

Para que possamos determinar quais variáveis realmente contribuem para o modelo final, precisamos de uma medida objetiva.

Para cada conjunto de $k = p - 1$ preditores, existem 2^{p-1} combinações de variáveis.

Para que possamos determinar quais variáveis realmente contribuem para o modelo final, precisamos de uma medida objetiva.

Já conhecemos uma dessas medidas, o ...

Para cada conjunto de $k = p - 1$ preditores, existem 2^{p-1} combinações de variáveis.

Para que possamos determinar quais variáveis realmente contribuem para o modelo final, precisamos de uma medida objetiva.

Já conhecemos uma dessas medidas, o ... $R^2_{ajustado}$.

Para cada conjunto de $k = p - 1$ preditores, existem 2^{p-1} combinações de variáveis.

Para que possamos determinar quais variáveis realmente contribuem para o modelo final, precisamos de uma medida objetiva.

Já conhecemos uma dessas medidas, o ... $R^2_{ajustado}$.

Podemos também usar os p-valores de cada coeficiente.

Para cada conjunto de $k = p - 1$ preditores, existem 2^{p-1} combinações de variáveis.

Para que possamos determinar quais variáveis realmente contribuem para o modelo final, precisamos de uma medida objetiva.

Já conhecemos uma dessas medidas, o ... $R^2_{ajustado}$.

Podemos também usar os p-valores de cada coeficiente.

Mas, por causa da multicolinearidade, os p-valores podem esconder variáveis importantes, mas correlacionadas.

A medida conhecida como AIC (Akaike's Information Criterion) é a mais comumente usada para comparação de modelos.

A medida conhecida como AIC (Akaike's Information Criterion) é a mais comumente usada para comparação de modelos.

O AIC é baseado no método de estimação por máxima verossimilhança, e na teoria da informação.

A medida conhecida como AIC (Akaike's Information Criterion) é a mais comumente usada para comparação de modelos.

O AIC é baseado no método de estimação por máxima verossimilhança, e na teoria da informação.

Definimos o AIC como:

$2k - 2\ln(L)$

k é o número de parâmetros.

L é a função de verossimilhança. Minimizamos essa função para encontrar a reta do modelo.

Para a comparação de modelos estimados por OLS, podemos usar a função equivalente:

$AIC = 2k + n \log(SQ_{Res}/\text{n})$

Para a comparação de modelos estimados por OLS, podemos usar a função equivalente:

$AIC = 2k + n\log(SQ_{Res}/\text{n})$

Interpretação:

Quanto melhor o ajuste do modelo, menor é $\ln(L)$ ou $n\log(SQ_{Res}/n)$

Para a comparação de modelos estimados por OLS, podemos usar a função equivalente:

$AIC = 2k + n\log(SQ_{Res}/\text{n})$

Interpretação:

Quanto melhor o ajuste do modelo, menor é $\ln(L)$ ou $n\log(SQ_{Res}/n)$

Mas quanto mais parâmetros adicionarmos, maior é $2k$

Para a comparação de modelos estimados por OLS, podemos usar a função equivalente:

$AIC = 2k + n\log(SQ_{Res}/\text{n})$

Interpretação:

Quanto melhor o ajuste do modelo, menor é $\ln(L)$ ou $n\log(SQ_{Res}/n)$

Mas quanto mais parâmetros adicionarmos, maior é $2k$

O melhor modelo minimiza o valor do AIC, através de uma combinação entre bom ajuste e **parcimônia**.

EXEMPLO: AIC

```r
m1 <- lm(qsec ~ hp, data=mtcars)
summary(m1)$r.squared
```

```
## [1] 0.5015804
```

```r
summary(m1)$adj.r.squared
```

```
## [1] 0.4849664
```

```r
AIC(m1)
```

```
## [1] 110.6665
```

```r
##
m2 <- lm(qsec ~ hp + wt, data=mtcars)
summary(m2)$r.squared
```

```
## [1] 0.6520291
```

```r
summary(m2)$adj.r.squared
```

```
## [1] 0.6280311
```

```r
AIC(m1,m2)
```

```
##    df      AIC
## m1  3 110.6665
## m2  4 101.1682
```

```r
m3 <- lm(qsec ~ hp + wt + disp, data=mtcars)
summary(m3)$r.squared
```

```
## [1] 0.6808292
```

```r
summary(m3)$adj.r.squared
```

```
## [1] 0.6466324
```

```r
AIC(m1,m2,m3)
```

```
##    df      AIC
## m1  3 110.6665
## m2  4 101.1682
## m3  5 100.4036
```

Outras medidas de comparação existem (Ex. AICc, BIC)

Diferentemente de testes de significância (*p-values* ou *likelihood tests*), os modelos comparados por AIC, AICc ou BIC não precisam ser aninhados (*nested*).

Contudo, a **variáveldependente e suas observações** precisam ser exatamente as mesmas. O que estamos avaliando e á capacidade de cada modelo em explicar/prever cada uma dessas variáveis.

Uma alternativa à escolha de um único modelo é o método de *model averaging*. Vários modelos são combinados, ponderados pela força da evidencia de cada um deles.

OBSERVAÇÕES IMPORTANTES

Outras medidas de comparação existem (Ex. AICc, BIC)

Diferentemente de testes de significância (*p-values* ou *likelihood tests*), os modelos comparados por AIC, AICc ou BIC não precisam ser aninhados (*nested*).

Contudo, a **variáveldependente e suas observações** precisam ser exatamente as mesmas. O que estamos avaliando e á capacidade de cada modelo em explicar/prever cada uma dessas variáveis.

Uma alternativa à escolha de um único modelo é o método de *model averaging*. Vários modelos são combinados, ponderados pela força da evidencia de cada um deles.

A melhor seleção de variáveis sempre dependerá da concordância do modelo com a teoria, e da aplicação esperada.

Quem vai receber o título: você ou o computador?

LEITURAS IMPORTANTES

Johnson JB, Omland KS (2004) Model selection in ecology and evolution. Trends in Ecology Evolution, 19, 101–8.

Whittingham MJ, Stephens P a, Bradbury RB, Freckleton RP (2006) Why do we still use stepwise modelling in ecology and behaviour? The Journal of Animal Ecology, 75, 1182–9.

Anderson D (2008) Model Based Inference in the Life Sciences: A Primer on Evidence. Springer New York, New York, NY, 146 pp.

Burnham KP (2004) Multimodel Inference: Understanding AIC and BIC in Model Selection. Sociological Methods Research, 33, 261–304.

Anderson D, Burnham KP (2000) Null hypothesis testing: problems, prevalence, and an alternative. The Journal of Wildlife Management, 64, 912–923.

Existem métodos automáticos de seleção de variáveis (ex. *stepwise*)

Tema da moda: **data mining**

Existem métodos automáticos de seleção de variáveis (ex. *stepwise*)

Tema da moda: **data mining**

Estes métodos podem oferecer contribuições importantes na **análise exploratória** do seu modelo.

Existem métodos automáticos de seleção de variáveis (ex. *stepwise*)

Tema da moda: **data mining**

Estes métodos podem oferecer contribuições importantes na **análise exploratória** do seu modelo.

Mas a melhor seleção final de variáveis sempre dependerá da concordância do modelo com a teoria, e da aplicação esperada.

Existem métodos automáticos de seleção de variáveis (ex. *stepwise*)

Tema da moda: **data mining**

Estes métodos podem oferecer contribuições importantes na **análise exploratória** do seu modelo.

Mas a melhor seleção final de variáveis sempre dependerá da concordância do modelo com a teoria, e da aplicação esperada.

Quem vai receber o diploma: voce ou o computador?

MÉTODO *STEPWISE*

Stepwise significa passo a passo.

Stepwise significa passo a passo.

O método pode ser aplicado de maneira crescente (*forward*) ou decrescente (*backward*).

Stepwise significa passo a passo.

O método pode ser aplicado de maneira crescente (*forward*) ou decrescente (*backward*).

No modo *forward*, começamos com uma única variável, e vamos progressivamente adicionando mais variáveis, testando o ganho em poder explicativo a cada nova adição.

MÉTODO *STEPWISE*

Stepwise significa passo a passo.

O método pode ser aplicado de maneira crescente (*forward*) ou decrescente (*backward*).

No modo *forward*, começamos com uma única variável, e vamos progressivamente adicionando mais variáveis, testando o ganho em poder explicativo a cada nova adição.

No modo *backward*, começamos com todas as variáveis, e vamos progressivamente eliminando cada uma, testando a perda em poder explicativo a cada nova adição.

O método usado para calcular esse ”ganho”ou ”perda”de informação pode variar de acordo com a escolha do usuário

O método usado para calcular esse ”ganho”ou ”perda”de informação pode variar de acordo com a escolha do usuário

O método stepwise original é baseado em testes F de significância, e bastante criticado por sua ampla sensitividade à multicolinearidade.

O método usado para calcular esse ”ganho”ou ”perda”de informação pode variar de acordo com a escolha do usuário

O método stepwise original é baseado em testes F de significância, e bastante criticado por sua ampla sensitividade à multicolinearidade.

O uso de medidas mais robustas como o AIC reduzem, mas não eliminam, esse problema.

O método usado para calcular esse "ganho"ou "perda"de informação pode variar de acordo com a escolha do usuário

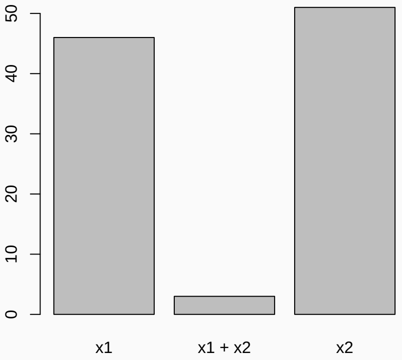
O método stepwise original é baseado em testes F de significância, e bastante criticado por sua ampla sensitividade à multicolinearidade.

O uso de medidas mais robustas como o AIC reduzem, mas não eliminam, esse problema.

Quando duas variáveis são muito parecidas, a escolha se torna arbitrária, e somente o suporte teórico (i.e. bom senso) pode resolver o problema.

MÉTODO *STEPWISE*

```r
form <- vector()

for(i in c(1:100)){
    x1 <- runif(30,0,20)
    x2 <- x1 + rnorm(30,0,1)
   y <- 3 + 2.3*x1 + 2.1*x2 + rnorm(30,0,10)
    m <- lm(y ~ x1 + x2)
    sm <- step(m, trace=0)
   form <- c(form, as.character(formula(sm))[3])
  }

barplot(table(factor(form)))
```



MÉTODO *STEPWISE*

Uma alternativa interessante pode ser combinar o método *stepwise* com métodos de aleatorização. Mas mesmo assim, continua sendo uma análise exploratória.

```
m1 <- lm (qsec ~ ., mtcars)

library(bootStepAIC)

## Loading required package:  MASS

m.boot <- boot.stepAIC(m1,mtcars,B = 10, direction="both")
```

```
m.boot$Covariates

##       (%)
## wt   100
## gear  90
## am    80
## vs    80
## cyl   70
## hp    70
## disp  60
## mpg   60
## carb  20
## drat  20
```

```
m.boot$Sign

##           + (%)     - (%)
## mpg  100.00000   0.00000
## vs   100.00000   0.00000
## wt   100.00000   0.00000
## hp    28.57143  71.42857
## am    12.50000  87.50000
## gear  11.11111  88.88889
## carb   0.00000 100.00000
## cyl    0.00000 100.00000
## disp   0.00000 100.00000
## drat   0.00000 100.00000
```

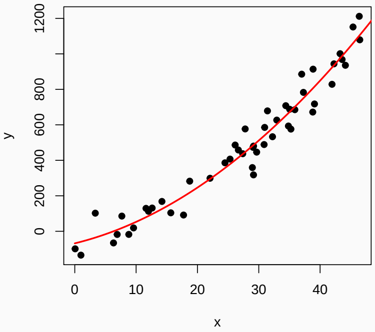
```
m.boot$Significance

##             (%)
## carb 100.00000
## wt   100.00000
## gear  88.88889
## vs    87.50000
## cyl   85.71429
## disp  83.33333
## am    75.00000
## mpg   66.66667
## hp    57.14286
## drat  50.00000
```

```
m.boot[4:6]

## $OrigModel
##
## Call:
## lm(formula = qsec ~ ., data = mtcars)
##
## Coefficients:
## (Intercept)          mpg          cyl         disp           hp         drat
##   17.776177     0.069048    -0.362678    -0.007501    -0.001563    -0.131064
##          wt           vs           am         gear         carb
##    1.496332     0.970035    -0.901186    -0.201285    -0.273598
##
##
## $OrigStepAIC
##
## Call:
## lm(formula = qsec ~ cyl + disp + wt + vs + am + carb, data = mtcars)
##
## Coefficients:
## (Intercept)          cyl         disp           wt           vs           am
##   18.611144    -0.369984    -0.008899     1.475086     0.968162    -0.902579
##        carb
##   -0.434722
##
##
## $direction
## [1] "both"
```

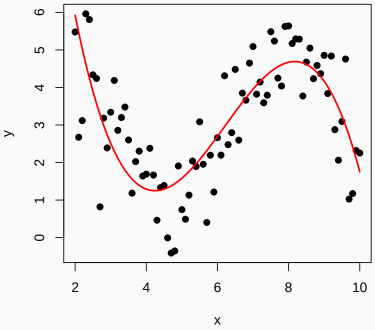
```
m.boot[7:8]

## $k
## [1] 2
##
## $BootStepAIC
## $BootStepAIC[[1]]
##
## Call:
## lm(formula = qsec ~ cyl + hp + wt + vs + gear, data = boot.data)
##
## Coefficients:
## (Intercept)          cyl           hp           wt           vs         gear
##    20.55994     -0.29160     -0.01034      0.68602      1.42131     -0.63289
##
##
## $BootStepAIC[[2]]
##
## Call:
## lm(formula = qsec ~ mpg + cyl + disp + wt + vs + am + gear, data = boot.data)
##
## Coefficients:
## (Intercept)          mpg          cyl         disp           wt           vs
##    19.66381      0.09939     -0.47900     -0.00653      1.02154      0.86578
##          am         gear
##    -1.35078     -0.65210
##
##
```

MAIS EXTENSÕES DOS MODELOS LINEARES GERAIS

```r
set.seed(234)

x <- runif(50,0,50)

y <- 5 + 1.3*x + 0.5*x^2 + rnorm(50,0,100)

m <- lm(y ~ x + I(x^2))

plot(x,y, pch=19)

xnovo <- data.frame( x = seq(0,50,by=0.5))

p <- predict(m,xnovo)

lines(xnovo[,1],p, lwd=2,col='red')
```

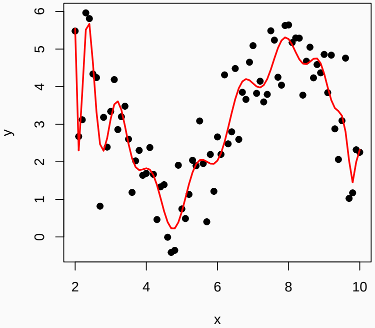


```r
set.seed(234)

x <- seq(2,10,by=0.1)

y <- 3 + 2*sin(x) + rnorm(81,0,1)

m <- lm(y ~ poly(x,3))

plot(x,y, pch=19)

xnovo <- data.frame( x = seq(2,10,by=0.1))

p <- predict(m,xnovo)

lines(xnovo[,1],p, lwd=2,col='red')
```



Cuidado com polinômios exagerados!

```
set.seed(234)

x <- seq(2,10,by=0.1)

y <- 3 + 2*sin(x) + rnorm(81,0,1)

m <- lm(y ~ poly(x,20))

plot(x,y, pch=19)

xnovo <- data.frame( x = seq(2,10,by=0.1))

p <- predict(m,xnovo)

lines(xnovo[,1],p, lwd=2,col='red')
```



Como saber se um modelo polinomial é melhor do que um modelo linear, e onde parar?

- Não podemos usar R^2, porque ele sempre aumenta...

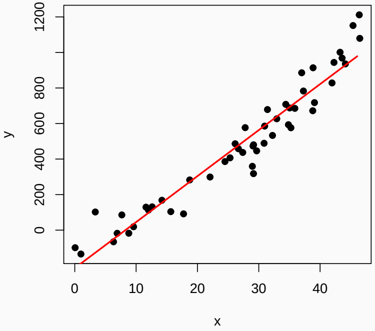
Como saber se um modelo polinomial é melhor do que um modelo linear, e onde parar?

- Não podemos usar R^2, porque ele sempre aumenta...
- Não podemos usar SQ_{reg}, porque os residuos ficam cada vez menores com mais termos

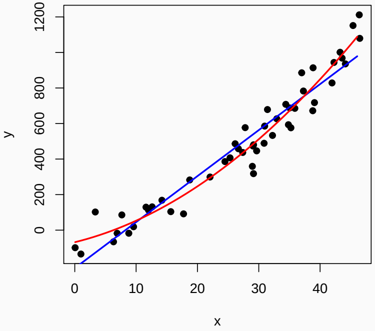
Como saber se um modelo polinomial é melhor do que um modelo linear, e onde parar?

- Não podemos usar R^2, porque ele sempre aumenta...
- Não podemos usar SQ_{reg}, porque os residuos ficam cada vez menores com mais termos
- Mas podemos usar o AIC!

```
set.seed(234)
x <- runif(50,0,50)
y <- 5 + 1.3*x + 0.5*x^2 + rnorm(50,0,100)

m1 <- lm(y ~ x)

AIC(m1)
## [1] 602.9227
```

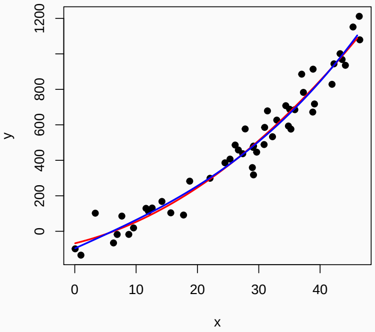


```
m2 <- lm(y ~ poly(x,2))

AIC(m1,m2)

##    df      AIC
## m1  3 602.9227
## m2  4 580.6495
```



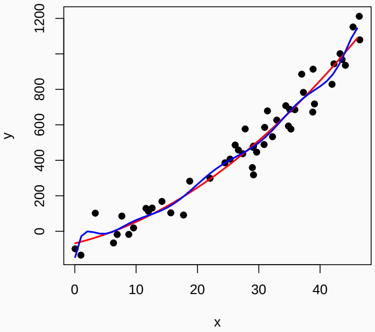
```r
m3 <- lm(y ~ poly(x,3))


AIC(m1,m2,m3)
```

```
##    df      AIC
## m1  3 602.9227
## m2  4 580.6495
## m3  5 581.8052
```



```
m10 <- lm(y ~poly(x,10))

AIC(m1,m2,m3,m10)
```

```
##     df      AIC
## m1   3 602.9227
## m2   4 580.6495
## m3   5 581.8052
## m10 12 592.0629
```



Limitações dos modelos polinomiais

- Os coeficientes não tem uma interpreteção direta

Limitações dos modelos polinomiais

- Os coeficientes não tem uma interpreteção direta
    - Não posso manter X_1 constante e variar X

Limitações dos modelos polinomiais

- Os coeficientes não tem uma interpreteção direta
  - Não posso manter X_1 constante e variar X
- Extrapolações são muito pouco confiáveis, pois o efeito dos termos de potência é muito forte fora do escopo

Para comparação de coeficientes com unidades diferentes

- Obtida através de normalização das variáveis:

Para comparação de coeficientes com unidades diferentes

- Obtida através de normalização das variáveis:
- $\dfrac{X - \bar{X}}{s}$

Para comparação de coeficientes com unidades diferentes

- Obtida através de normalização das variáveis:
- $\frac{X - \bar{X}}{s}$
- A subtração da média é chamada de **centralização** dos dados

Para comparação de coeficientes com unidades diferentes

- Obtida através de normalização das variáveis:
- $\dfrac{X - \bar{X}}{s}$
- A subtração da média é chamada de **centralização** dos dados
- A divisão pelo desvio padrão fornece a **normalização** da variância

Para comparação de coeficientes com unidades diferentes

- Obtida através de normalização das variáveis:
- $\dfrac{X - \bar{X}}{s}$
- A subtração da média é chamada de **centralização** dos dados
- A divisão pelo desvio padrão fornece a **normalização** da variância
- Bônus: A estimativa de β_0 se torna o valor de $E(Y)$ para os valores médios de X

REGRESSÃO NORMALIZADA

```r
x1 <- runif(20,0,100)
x2 <- runif(20,0,10)
y <- 5 + 3*x1 + 30*x2 + rnorm(20,0,20)
m <- lm(y ~ x1 + x2)
summary(m)
```

```
##
## Call:
## lm(formula = y ~ x1 + x2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -49.956  -9.722   5.638  12.781  27.107
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.2083    12.2693    0.18    0.859
## x1            3.1714     0.1787   17.75 2.09e-12 ***
## x2           28.0824     1.8342   15.31 2.24e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05
##
## Residual standard error: 22.95 on 17 degrees of
## Multiple R-squared:  0.9741,^^IAdjusted R-squa
## F-statistic: 320.1 on 2 and 17 DF,  p-value: 3.2
```

```r
x1n <- (x1-mean(x1))/sd(x1)
x2n <- (x2-mean(x2))/sd(x2)
mn <- lm(y ~ x1n + x2n)
summary(mn)
```

```
##
## Call:
## lm(formula = y ~ x1n + x2n)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -49.956  -9.722   5.638  12.781  27.107
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  280.077      5.131   54.59  < 2e-16 **
## x1n           94.397      5.319   17.75 2.09e-12 ***
## x2n           81.430      5.319   15.31 2.24e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05
##
## Residual standard error: 22.95 on 17 degrees of
## Multiple R-squared:  0.9741,^^IAdjusted R-squar
## F-statistic: 320.1 on 2 and 17 DF,  p-value: 3.2
```

MODELOS LINEARES GENERALIZADOS

A **Regressão Logística** inverte o jogo, e prediz uma variável qualitativa usando dados contínuos

- Exemplo: Qual a probabilidade de um aluno ser aceito em um programa de pós-graduação, dada a sua média escolar, nota geral das provas, e avaliação das cartas de recomendação?
- O "truque" é transformar um dado qualitativo (aprovado/não-aprovado) em um dado quantitativo

Probabilidades:

- Aprovado/Não-Aprovado $\sim$ Binomial(n, p)
- Aprovação possui uma probabilidade p
- Reprovação possui probabilidade $1 - p$

Probabilidades:

- Aprovado/Não-Aprovado $\sim$ Binomial(n, p)
- Aprovação possui uma probabilidade p
- Reprovação possui probabilidade $1 - p$
- Probabilidades possuem valores contínuos, mas …

Probabilidades:

- Aprovado/Não-Aprovado $\sim$ Binomial(n, p)
- Aprovação possui uma probabilidade p
- Reprovação possui probabilidade $1 - p$
- Probabilidades possuem valores contínuos, mas ...possuem limites (0,1), que dificultam a modelagem

Probabilidades:

- Aprovado/Não-Aprovado $\sim$ Binomial(n, p)
- Aprovação possui uma probabilidade p
- Reprovação possui probabilidade $1 - p$
- Probabilidades possuem valores contínuos, mas …possuem limites (0,1), que dificultam a modelagem
- Quanto maiores as probabilidades, maior a dificuldade para aumentá-las (i.e. não-linear)

Chances (ou possibilidades, *odds*)

- Razão entre probabilidades
- Para uma distribuição binomial, $o = \dfrac{p}{1-p}$

Chances (ou possibilidades, *odds*)

- Razão entre probabilidades
- Para uma distribuição binomial, $o = \dfrac{p}{1-p}$
- Ex.: Se $p = 0.8$, $o = \dfrac{0.8}{0.2} = 4$
- A chance é de 4 para 1

Chances (ou possibilidades, *odds*)

- Razão entre probabilidades
- Para uma distribuição binomial, $o = \dfrac{p}{1-p}$
- Ex.: Se $p = 0.8$, $o = \dfrac{0.8}{0.2} = 4$
- A chance é de 4 para 1
- Se $p = 0$, $o = 0$; se $p = 1$, $o = Inf$ (por limite)

Chances (ou possibilidades, *odds*)

- Razão entre probabilidades
- Para uma distribuição binomial, $o = \dfrac{p}{1-p}$
- Ex.: Se $p = 0.8$, $o = \dfrac{0.8}{0.2} = 4$
- A chance é de 4 para 1
- Se $p = 0$, $o = 0$; se $p = 1$, $o = Inf$ (por limite)
- Resolvemos metade do problema dos limites $(0, 1)$.

Log das Chances (*log odds*): transformação logito (*logit*)

- logaritmo da chance
- Se $p = 0$, $o = 0$ e $\log(0) = -Inf$ (por limite)
- Se $p = 1$, $o = Inf$ e $log(Inf) = Inf$ (por limite)

Log das Chances (*log odds*): transformação logito (*logit*)

- logaritmo da chance
- Se $p = 0$, $o = 0$ e $\log(0) = -Inf$ (por limite)
- Se $p = 1$, $o = Inf$ e $log(Inf) = Inf$ (por limite)
- Agora podemos escrever $\log\left(\frac{p(X)}{1-p(X)}\right) = \beta_0 + \beta_1 X + \varepsilon$

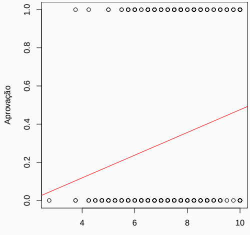
Log das Chances (*log odds*): transformação logito (*logit*)

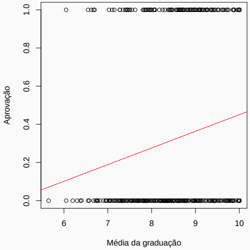
- logaritmo da chance
- Se $p = 0$, $o = 0$ e $\log(0) = -Inf$ (por limite)
- Se $p = 1$, $o = Inf$ e $log(Inf) = Inf$ (por limite)
- Agora podemos escrever $\log\left(\frac{p(X)}{1 - p(X)}\right) = \beta_0 + \beta_1 X + \varepsilon$
- Equivale a $p(X) = \frac{e^{\beta_0 + \beta_1 X}}{e^{\beta_0 + \beta_1 X} + 1} + \varepsilon = \frac{1}{1 - e^{\beta_0 + \beta_1 X}} + \varepsilon$

```r
phd <- read.csv('inpe_aprov_fake.csv')
head(phd)
```

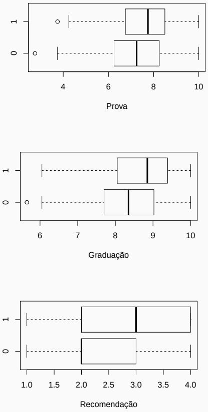
```
##   X aprovado prova   grad reco
## 1 1        0  4.75  9.025    2
## 2 2        1  8.25  9.175    2
## 3 3        1 10.00 10.000    4
## 4 4        1  8.00  7.975    1
## 5 5        0  6.50  7.325    1
## 6 6        1  9.50  7.500    3
```

```r
par(mar=c(4,4,0,0))
plot(aprovado ~ prova, data=phd)
plot(aprovado ~ grad, data=phd)
```



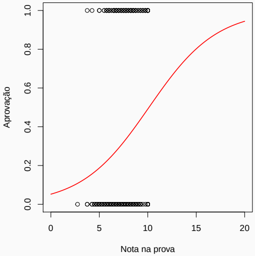


```
phd <- read.csv('inpe_aprov_fake.csv')
head(phd)
par(mar=c(7,2,1,0))
boxplot(prova ~ fac-
tor(aprovado), data=phd,xlab="Prova",horizontal=TRUE)
boxplot(grad ~fac-
tor(aprovado), data=phd,xlab="Graduação", horizontal=TRUE)
boxplot(reco ~ fac-
tor(aprovado), data=phd,horizontal=TRUE,xlab="Recomendação")
```



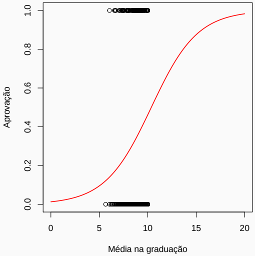
```r
m <- glm(aprovado ~ prova, phd, family='binomial')
summary(m)
```

```
##
## Call:
## glm(formula = aprovado ~ prova, family = "binomial", data = phd)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.1623  -0.9052  -0.7547   1.3486   1.9879
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.90134    0.60604  -4.787 1.69e-06 ***
## prova        0.28658    0.07888   3.633  0.00028 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 499.98  on 399  degrees of freedom
## Residual deviance: 486.06  on 398  degrees of freedom
## AIC: 490.06
##
## Number of Fisher Scoring iterations: 4
```



```r
m <- glm(aprovado ~ grad, phd, family='binomial')
summary(m)
```

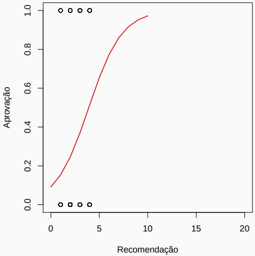
```
##
## Call:
## glm(formula = aprovado ~ grad, family = "binomial", data = phd)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.1131  -0.8874  -0.7566   1.3305   1.9824
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -4.3576     1.0353  -4.209 2.57e-05 ***
## grad          0.4204     0.1195   3.517 0.000437 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 499.98  on 399  degrees of freedom
## Residual deviance: 486.97  on 398  degrees of freedom
## AIC: 490.97
##
## Number of Fisher Scoring iterations: 4
```



REGRESSÃO LOGÍSTICA

```
m <- glm(aprovado ~ reco, phd, family='binomial')
summary(m)
```

```
##
## Call:
## glm(formula = aprovado ~ reco, family = "binomial", data = phd)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.1989  -0.9599  -0.7508   1.1561   1.9365
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -2.2948     0.3524  -6.511 7.46e-11 ***
## reco          0.5863     0.1240   4.728 2.26e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 499.98  on 399  degrees of freedom
## Residual deviance: 475.71  on 398  degrees of freedom
## AIC: 479.71
##
## Number of Fisher Scoring iterations: 4
```



REGRESSÃO LOGÍSTICA

```r
m <- glm(aprovado ~ grad + prova + reco, phd, family='binomial')
summary(m)
```

```
##
## Call:
## glm(formula = aprovado ~ grad + prova + reco, family = "binomial",
##     data = phd)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.5802  -0.8848  -0.6382   1.1575   2.1732
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -6.24971    1.15573  -5.408 6.39e-08 ***
## grad         0.31081    0.13099   2.373   0.0177 *
## prova        0.18352    0.08735   2.101   0.0356 *
## reco         0.56003    0.12714   4.405 1.06e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 499.98  on 399  degrees of freedom
## Residual deviance: 459.44  on 396  degrees of freedom
## AIC: 467.44
##
## Number of Fisher Scoring iterations: 4
```

```r
# exponencial dos coeficientes = odds
exp(m$coefficients[2:4])
```

```
##     grad    prova     reco
## 1.364524 1.201435 1.750727
```

A regressão logística é um caso específico dos chamados **Modelos Lineares *Generalizados***

- Estendem os modelos lineares gerais para dados não-normais
- Utilizam a chamada *função link* para relacionar os dados originais com uma distribuição normal

Função Link	Distribuições
Identidade	Normal
Inversa	Exponencial, Gamma
Log	Poisson
Logito	Bernoulli, Binomial, Multinomial

Dúvida: usar uma função link não é a mesma coisa que transformar a variável?

Dúvida: usar uma função link não é a mesma coisa que transformar a variável?

- Não exatamente. A função link transforma $E(Y)$, e não Y

Dúvida: usar uma função link não é a mesma coisa que transformar a variável?

- Não exatamente. A função link transforma $E(Y)$, e não Y
- Por exemplo: Função Log
  - Se você transforma a variável, está estimando $E(log(Y))$

Dúvida: usar uma função link não é a mesma coisa que transformar a variável?

- Não exatamente. A função link transforma $E(Y)$, e não Y
- Por exemplo: Função Log
  - Se você transforma a variável, está estimando $E(log(Y))$
  - Um GLM com função link estima $log(E(Y))$

Dúvida: usar uma função link não é a mesma coisa que transformar a variável?

- Não exatamente. A função link transforma $E(Y)$, e não Y
- Por exemplo: Função Log
  - Se você transforma a variável, está estimando $E(log(Y))$
  - Um GLM com função link estima $log(E(Y))$
- Os parâmetros serão diferentes, e o GLM produz estimativas mais precisas

MÍNIMOS QUADRADOS GENERALIZADOS

Método generalizado de mínimos quadrados para o ajuste de modelos

- Generalized Least Squares (GLS) $\neq$ Generalized Linear Model (GLM)
- Permite estimar parâmetros quando os resíduos são heteroscedásticos
- Permite estimar parâmetros quando os resíduos são correlacionados

Na regressão por mínimos quadrados ordinários (OLS), assumimos que $e \sim N(0, s^2)$

Como poderíamos modificar a descrição de e, para dados onde a variância cresce com os valores de X?

Na regressão por mínimos quadrados ordinários (OLS), assumimos que $e \sim N(0, s^2)$

Como poderíamos modificar a descrição de e, para dados onde a variância cresce com os valores de X?

$e \sim N(0, s^2 \times X)$: modelo fixo de estrutura de variância

Na regressão por mínimos quadrados ordinários (OLS), assumimos que $e \sim N(0, s^2)$

Como poderíamos modificar a descrição de e, para dados onde a variância cresce com os valores de X?

$e \sim N(0, s^2 \times X)$: modelo fixo de estrutura de variância

Poderiam haver outros tipos de relação entre $Var(e)$ e X?

Poderiam haver outros tipos de relação entre $Var(e)$ e X?

$e \sim N(0, s_j^2)$: modelo VarIdent de estrutura de variânica (X categórico)

Poderiam haver outros tipos de relação entre $Var(e)$ e X?

$e \sim N(0, s_j^2)$: modelo VarIdent de estrutura de variânica (X categórico)

$e \sim N(0, s^2 \times |X_j|^{2\delta})$: modelo VarPower (não funciona se X tem zeros)

Poderiam haver outros tipos de relação entre $Var(e)$ e X?

$e \sim N(0, s_j^2)$: modelo VarIdent de estrutura de variânica (X categórico)

$e \sim N(0, s^2 \times |X_j|^{2\delta})$: modelo VarPower (não funciona se X tem zeros)

$e \sim N(0, s^2 \times e^{2\delta \times X})$: modelo VarExp (funciona se X tem zeros)

Poderiam haver outros tipos de relação entre $Var(e)$ e X?

$e \sim N(0, s_j^2)$: modelo VarIdent de estrutura de variânica (X categórico)

$e \sim N(0, s^2 \times |X_j|^{2\delta})$: modelo VarPower (não funciona se X tem zeros)

$e \sim N(0, s^2 \times e^{2\delta \times X})$: modelo VarExp (funciona se X tem zeros)

$e \sim N(0, s^2 \times (\delta_1 \times |X|^{2\delta_2})^2)$: modelo VarConstPower

Poderiam haver outros tipos de relação entre $Var(e)$ e X?

$e \sim N(0, s_j^2)$: modelo VarIdent de estrutura de variânica (X categórico)

$e \sim N(0, s^2 \times |X_j|^{2\delta})$: modelo VarPower (não funciona se X tem zeros)

$e \sim N(0, s^2 \times e^{2\delta \times X})$: modelo VarExp (funciona se X tem zeros)

$e \sim N(0, s^2 \times (\delta_1 \times |X|^{2\delta_2})^2)$: modelo VarConstPower

$e \sim N(0, s_j^2 \times e^{2\delta_2 \times X})$: modelo VarComb

Poderiam haver outros tipos de relação entre $Var(e)$ e X?

$e \sim N(0, s_j^2)$: modelo VarIdent de estrutura de variânica (X categórico)

$e \sim N(0, s^2 \times |X_j|^{2\delta})$: modelo VarPower (não funciona se X tem zeros)

$e \sim N(0, s^2 \times e^{2\delta \times X})$: modelo VarExp (funciona se X tem zeros)

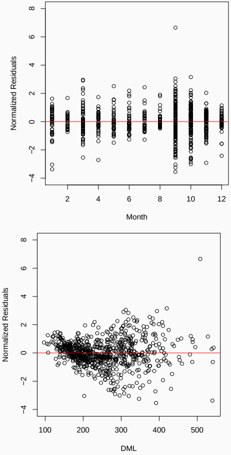
$e \sim N(0, s^2 \times (\delta_1 \times |X|^{2\delta_2})^2)$: modelo VarConstPower

$e \sim N(0, s_j^2 \times e^{2\delta_2 \times X})$: modelo VarComb

Para saber mais: Zuur et al. (2009). Mixed effects models and extensions in Ecology with R. Springer.

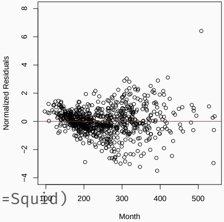
```
# Exemplo de Zuur et al. (2009)m - Lulas <-
library(nlme)
Squid<-read.table(file="Squid.txt",header=TRUE)
Squid$fMONTH=factor(Squid$MONTH)
M1 <- lm(Testisweight ~ DML * fMONTH,data=Squid)
anova(M1)
```

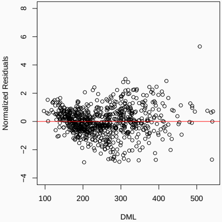
```
## Analysis of Variance Table
##
## Response: Testisweight
##              Df  Sum Sq Mean Sq  F value    Pr(>F)
## DML           1 11247.2 11247.2 1732.082 < 2.2e-16 ***
## fMONTH       11  2099.1   190.8   29.388 < 2.2e-16 ***
## DML:fMONTH   11  1678.0   152.5   23.492 < 2.2e-16 ***
## Residuals   744  4831.1     6.5
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



```
# Usando VarIdent para corrigir o efeito de DML
M.lm<-gls(Testisweight~DML*fMONTH,data=Squid)
M.gls1<-gls(Testisweight~DML*fMONTH,weights=varFixed(~DML),data=Squid)
AIC(M.lm,M.gls1)
```

```
##        df      AIC
## M.lm   25 3752.084
## M.gls1 25 3620.898
```

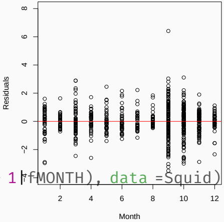


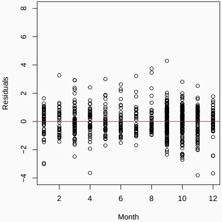


GLS

```
# Usando VarIdent para corrigir o efeito de DML
M.gls2 <- gls(Testisweight ~ DML*fMONTH, weights=varIdent(form= ~ 1|fMONTH), data =Squid)
AIC(M.lm,M.gls1,M.gls2)
```

```
##        df      AIC
## M.lm   25 3752.084
## M.gls1 25 3620.898
## M.gls2 36 3614.436
```

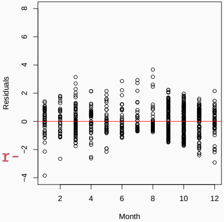


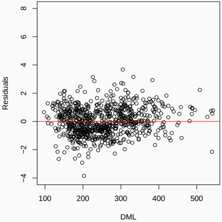


GLS

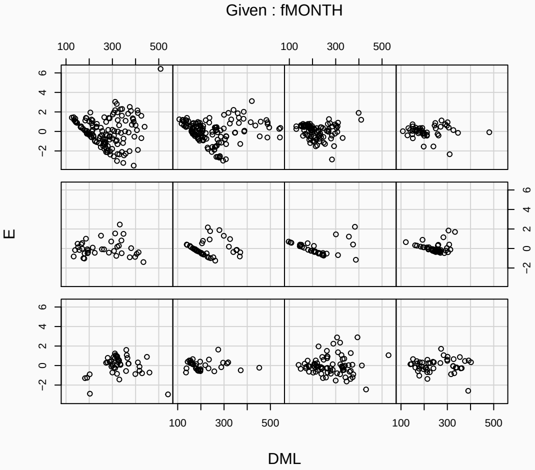
```
# Usando VarPower para corrigir o efeito de DML e MONTH
M.gls3<-gls(Testisweight ~ DML * fMONTH, data = Squid,weights = var-
Power(form=~ DML | fMONTH))
AIC(M.lm,M.gls1,M.gls2,M.gls3)
```

```
##         df      AIC
## M.lm    25 3752.084
## M.gls1  25 3620.898
## M.gls2  36 3614.436
## M.gls3  37 3407.511
```

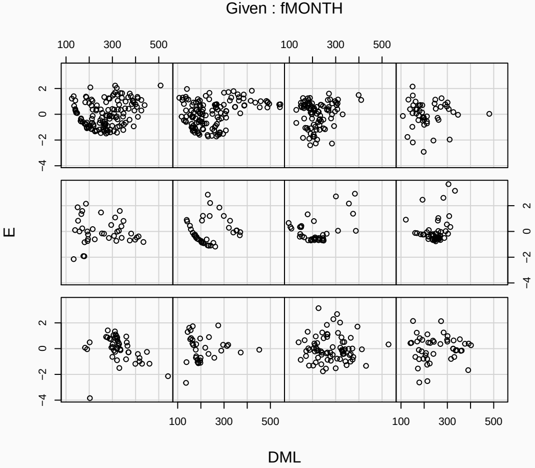




Resíduos no modelo estimado por OLS



GLS com varPower(form= DML | fMONTH)



Como poderíamos modificar a descrição de e, para dados onde os valores são correlacionados?

Como poderíamos modificar a descrição de e, para dados onde os valores são correlacionados?

$e \sim N(0, \rho \times s^2)$: modelo de correlação simétrico

ρ é a correlação entre X_i e $X_{j \neq i}$

Como poderíamos modificar a descrição de e, para dados onde os valores são correlacionados?

$e \sim N(0, \rho \times s^2)$: modelo de correlação simétrico

ρ é a correlação entre X_i e $X_{j \neq i}$

$e \sim N(0, \rho \times s^2)$: modelo de correlação autoregressivo (AR) de ordem 1

cor(X_i,X_{i-1}) = ρ; cor(X_i,X_{i-2}) = ρ^2; cor(X_i,X_{i-n}) = ρ^n

Como poderíamos modificar a descrição de e, para dados onde os valores são correlacionados?

$e \sim N(0, \rho \times s^2)$: modelo de correlação simétrico

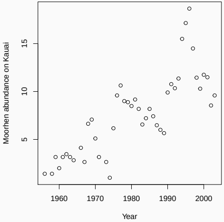
ρ é a correlação entre X_i e $X_{j \neq i}$

$e \sim N(0, \rho \times s^2)$: modelo de correlação autoregressivo (AR) de ordem 1

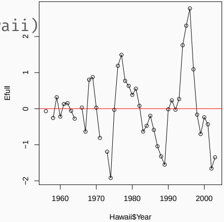
cor(X_i,X_{i-1}) = ρ; cor(X_i,X_{i-2}) = ρ^2; cor(X_i,X_{i-n}) = ρ^n

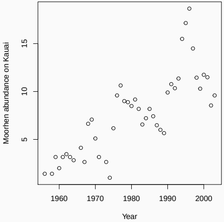
Existem modelos (bem) mais complexos.

GLS

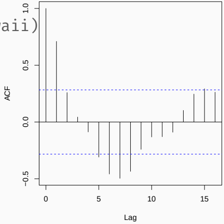



```r
# Exemplo de Zuur et al. (2009)m - Pássaros
Hawaii<-read.table(file="Hawaii.txt",header=TRUE)
Hawaii$Birds<-sqrt(Hawaii$Moorhen.Kauai)
M0 <- gls(Birds ~ Rainfall + Year, na.action = na.omit, data = Hawaii)
```





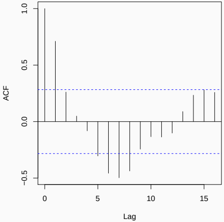

```
# Exemplo de Zuur et al. (2009)m - Pássaros
Hawaii<-read.table(file="Hawaii.txt",header=TRUE)
Hawaii$Birds<-sqrt(Hawaii$Moorhen.Kauai)
M0 <- gls(Birds ~ Rainfall + Year, na.action = na.omit, data = Hawaii)
```

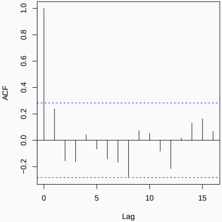


GLS

```r
# Exemplo de Zuur et al. (2009) - Pássaros
M1<-gls(Birds ~ Rainfall + Year, na.action = na.omit,
        correlation = corCompSymm(form =~ Year),
        data=Hawaii)

M2<-gls(Birds ~ Rainfall + Year, na.action = na.omit,
      correlation = corAR1(form =~ Year), data = Hawaii)

AIC(M0,M1,M2)
```

```
##    df      AIC
## M0  4 228.4798
## M1  5 230.4798
## M2  5 199.1394
```





Regressão Robusta

Regressão por Quantil

Regressão Não-Linear

Modelos Aditivos Generalizados (Generalized Additive Models, GAM)

Splines

Mixed Models

LEITURAS RECOMENDADAS:

Bolker BM, Brooks ME, Clark CJ, Geange SW, Poulsen JR, Stevens MHH, White J-SS (2009) Generalized linear mixed models: a practical guide for ecology and evolution. Trends in ecology evolution, 24, 127–35.

Yee TW, Mitchell ND (1991) Generalized additive models in plant ecology. Journal of Vegetation Science, 2, 587–602.

Zuur A, Ieno E, Walker N, Saveliev A, Smith G (2009) Mixed Effects Models and Extensions in Ecology with R. Springer.

Pinheiro J, Bates D (2000) Mixed-Effects Models in S and S-PLUS. Springer.