# Modelos Lineares Generalizados - Verificação do Ajuste do Modelo

Erica Castilho Rodrigues

9 de Abril de 2015

# Funï¿$\frac{1}{2}$ï¿$\frac{1}{2}$o Deviance

# Função *Deviance*

- Podemos ver o ajuste de um modelo a um conjunto de dados como:
  - uma forma de substituir **y** por $\hat{\boldsymbol{\mu}}$
- $\hat{\boldsymbol{\mu}}$ é o vetor de médias estimado a partir do modelo.
- Em Regressão Linear, substuíamos **y** por $\hat{\mathbf{y}}$.
- Objetivo: encontrar um modelo com poucos parâmetros tal que essa substuitção seja razoável.
- Os $\hat{\mu}$'s não serão exatamente iguais aos $y$'s.
- O quanto eles diferem?
- Pequena discrepância - tolerável.
- Discrepância grande - o modelo não está bom.

- Suponha que já encontramos:
  - qual a distribuição adequada;
  - qual a função de ligação mais razoável.
- Temos então

$$Y_i \sim \text{família exponencial canônica} \quad E(Y_i) = \mu_i = g(\mathbf{x}_i^T \boldsymbol{\beta}) \, .$$

- Objetivo: determinar quantos termos são necessários no preditor linear

$$\mathbf{x}_i^T \boldsymbol{\beta} = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip} \, .$$

- Se optamos por um número grande de termos:
  - o modelo vai explicar bem os dados;
  - vai ficar complexo e com muitos parâmetros a estimar.

- Se optamos por um modelo com poucas variáveis:
  - o modelo é simples e fácil de estimar;
  - provavelmente não se ajustará bem aos dados.
- Queremos encontrar um modelo intermediário.
- Com $n$ observações qual o número máximo de parâmetros que podemos ter? $n$.
- Caso contrário, não temos informação suficiente para estimar os parâmetros.
- Para encontrarmos uma reta precisamos de pelo menos dois pontos.

- Vejamos as nomenclaturas dos modelos.
- **Modelo Nulo:**
  - tem um único parâmetro ($\mu$);
  - esse parâmetro é comum a todos os dados;
  - não tem covariáveis;

$$\eta_i = \mu \,.$$

- Os $y$'s variam entre si apenas por causa do componente aleatório.
- Não variam devido aos diferentes níveis das covariávies.

- **Modelo Saturado ou Modelo Completo:**
  - tem o número máximo de parâmetros possível ($n$);
  - um parâmetro por observação;
  - não existe erro aleatório.
- Toda a variação é devida ao componente sistemático;
- Os $y$'s variam por causa dos níveis diferentes das covariáveis;
- O modelo explica toda variabilidade dos dados.

$$\eta_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_n X_{in}$$

onde $n$ é o tamanho da amostra.

- O modelo nulo é simples demais.
- O modelo saturado é não informativo, apenas repete os dados.
- No modelo saturado podemos incluir ainda termos de interação.
- Queremos comparar os modelos.
- Encontrar um, que esteja entre o nulo e o saturado, e seja adequado aos dados.
- Vamos comparar modelos encaixados:
  - um está contido no outro;
  - podemos ir de um para outro apenas acrescentando ou excluindo covariáveis.

- Chamamos o modelo intermediário de *p* parâmetros modelo corrente ou modelo sob pesquisa.
- Devemos ou não incluir uma variável no modelo?
- Esse parâmetro extra será útil?
- Se não incluimos o modelo fica mal ajustado?
- Podemos usar uma medida para avaliar se o modelo está bem ajustado: **Deviance** ou Função Desvio.

## Deviance ou Função Desvio

- Mede a qualidade de ajuste do modelo.
- É calculada como

$$D = 2[l(\mathbf{b}_{max}; \mathbf{y}) - l(\mathbf{b}, \mathbf{y})]$$

  onde
  - $l(\mathbf{b}_{max}; \mathbf{y})$ é o máximo da log-verossimilhança para o modelo saturado;
  - $l(\mathbf{b}, \mathbf{y})$ é o máximo da log-verossimilhança para o modelo sob pesquisa.
- Consiste em um teste da razão de verossimilhança que compara:
  - o modelo saturado com o modelo corrente.

- $D$ é **grande** se:
  - o ajuste o modelo corrente é muito **pior** do que do saturado.
- $D$ é **pequeno** se:
  - o ajuste do modelo corrente não é tão pior do que do saturado.
- O melhor ajuste possível é obtido com o modelo saturado.
- Pois ele explica toda a variablidade dos dados.

- Observe que essa estatística pode ser escrita como

$$\lambda = \log\left(\frac{f(\mathbf{b}_{max};\mathbf{y})}{f(\mathbf{b},\mathbf{y})}\right).$$

- O que isso te lembra?
- Teste da Razão de Verossimilhança.
- Estamos no fundo fazendo um Teste da Razão.
- Pode-se mostrar que a distribuição aproximada da estatística é

$$D \sim \chi^2_{(n-p)}$$

onde

  - $n$ é o número de observações na amostra;
  - $p$ é o número de parâmetros do modelo.

## Exemplo:

- Considere uma amostra aleatória

$$Y_1, Y_2, \ldots, Y_m$$

tal que $Y_i \sim Bin(n_i, \pi_i)$.
- Vamos encontrar a função Deviance para esse caso.
- A função de verossimlhança é dada por

$$f(\mathbf{y}, \pi_i) = \prod_i \binom{n_i}{y_i} \pi_i^{y_i} (1 - \pi_i)^{n_i - y_i}$$

tirando o logaritmo

$$l(\mathbf{y}, \pi_i) = \sum_i \left( \log \binom{n_i}{y_i} + y_i \log(\pi_i) + (n_i - y_i) \log(1 - \pi_i) \right)$$

$$= \sum_i \left( y_i \log \left( \frac{\pi_i}{1 - \pi_i} \right) + n_i \log(1 - \pi_i) + \log \binom{n_i}{y_i} \right)$$

**Exemplo: (continuação)**

- Vamos encontrar a log-verossimilhança do Modelo Saturado.
- No modelo saturado o valor predito é igual ao observado.
- O modelo não tem erro nenhum, se ajusta perfeitamente aos dados.
- Temos então que

$$(E(\hat{Y}_i)) = Y_i$$

como $E(Y_i) = n_i\pi_i$ temos que

$$(\hat{E}(Y_i)) = y_i = \hat{\pi}_i n_i \Rightarrow \hat{\pi}_i = \frac{y_i}{n_i}\ .$$

**Exemplo: (continuação)**

- Para obtermos o ponto onde a log-verossimilhança basta:
  - substituirmos o valor estimado $\hat{\pi}_i$ na função

$$l(\mathbf{b}_{max}, \mathbf{y}) = \sum_i \left( y_i \log\left(\frac{y_i/n_i}{1 - y_i/n_i}\right) + n_i \log(1 - y_i/n_i) + \log\binom{n_i}{y_i}\right)$$

## Exemplo: (continuação)

- Vamos agora calcular a log-verossimilhança máxima para o modelo sob análise.
- Vamos denotar por $\hat{\pi}_i$ as estimativas para os parâmetros $\pi_i$ obtidas através do modelo.
- Os valores $y_i$ são preditos da seguinte maneira

$$\hat{y}_i = n_i \hat{\pi}_i$$

  onde $\hat{\pi}_i$ são os valores estimados a partir do modelo sob análise.
- A log-verossimilhança avaliada nesse ponto é dada por

$$\begin{aligned} l(\mathbf{b}, \mathbf{y}) &= \sum_i \left( y_i \log\left(\frac{\hat{y}_i/n_i}{1-\hat{y}_i/n_i}\right) + n_i \log(1-\hat{y}_i/n_i) + \log\binom{n_i}{y_i}\right) \\ &= \sum_i \left( y_i \log\left(\frac{\hat{y}_i/n_i}{\frac{n_i-\hat{y}_i}{n_i}}\right) + n_i \log(\frac{n_i-\hat{y}_i}{n_i}) + \log\binom{n_i}{y_i}\right) \\ &= \sum_i \left( y_i \log\left(\frac{\hat{y}_i}{n_i-\hat{y}_i}\right) + n_i \log(\frac{n_i-\hat{y}_i}{n_i}) + \log\binom{n_i}{y_i}\right) \end{aligned}$$

## **Exemplo:** (continuação)

- Vamos agora subtrair as duas funções para obter a Deviance

$$D = 2[l(\mathbf{b}_{max};\mathbf{y}) - l(\mathbf{b},\mathbf{y})]$$

$$D = 2\left[\sum_i \left(y_i \log\left(\frac{y_i/n_i}{1-y_i/n_i}\right) + n_i \log(1-y_i/n_i) + \log\binom{n_i}{y_i}\right) - \right.$$

$$\left.\sum_i \left(y_i \log\left(\frac{\hat{y}_i}{n_i-\hat{y}_i}\right) + n_i \log(\frac{n_i-\hat{y}_i}{n_i}) + \log\binom{n_i}{y_i}\right)\right]$$

$$= 2\left[\sum_i y_i \log\left(\frac{y_i}{\hat{y}_i}\right) - y_i \log\left(\frac{n_i-\hat{y}_i}{n_i-y_i}\right) + n_i \log\left(\frac{n_i-\hat{y}_i}{n_i-y_i}\right)\right]$$

**Exemplo: (continuação)**

- Portanto

$$D = 2\left[\sum_i y_i \log\left(\frac{y_i}{\hat{y}_i}\right) + (n_i - y_i)\log\left(\frac{n_i - \hat{y}_i}{n_i - y_i}\right)\right]$$

é a função Deviance quando a veriável resposta tem distribuição Binomial.

**Exemplo:**

- Considere uma amostra aleatória

$$Y_1, Y_2, \ldots, Y_n$$

  são tais que $Y_i \sim N(\mu_i, \sigma^2)$.
- A média é modelada por

$$E(Y_i) = \mu_i = \mathbf{x}_i^T \beta \, .$$

- A função de ligação? Identidade.
- Vamos encontrar a função Deviance nesse caso.

## Exemplo: (continuação)

- A função de verossimilhança é dada por

$$f(\mathbf{y};\beta) = \prod_i \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y_i-\mu_i)^2} = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{1}{2\sigma^2}\sum_i (y_i-\mu_i)^2}$$

a log-verossimilhança é dada por

$$\log(f(\mathbf{y};\beta)) = -\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu_i)^2 - \frac{n}{2}\log(2\pi\sigma^2)\,.$$

**Exemplo: (continuação)**

- Para o modelo saturado temos que a média estimada pelo modelo é exatamente igual a observação.
- Portanto
$$\hat{E}(Y_i) = \hat{\mu}_i = y_i \,.$$
- Para obtermos o ponto onde a log-verossimilhança basta:
  - substituirmos o valor estimado $\hat{\pi}_i$ na log-verossimilhança

$$\begin{aligned}\log(f(\mathbf{y};\beta)) &= -\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - y_i)^2 - \frac{n}{2}\log(2\pi\sigma^2)\\ &= -\frac{n}{2}\log(2\pi\sigma^2)\,.\end{aligned}$$

**Exemplo: (continuação)**

- Para o modelo sob análise temos que

$$\hat{E}(Y_i) = \hat{\mu}_i = \mathbf{x}_i^T \hat{\beta}$$

onde $\hat{\beta}$ e o estimador de $\beta$.

- A log-verossimilhança do modelo sob análise fica então

$$\log(f(\mathbf{y};\beta)) = -\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mathbf{x}_i^T\hat{\beta})^2 - \frac{n}{2}\log(2\pi\sigma^2)$$

## **Exemplo:** (continuação)

- Temos então que

$$l(\mathbf{b}_{max}, \mathbf{y}) = -\frac{n}{2}\log(2\pi\sigma^2)$$

$$l(\mathbf{b}, \mathbf{y}) = -\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mathbf{x}_i^T\hat{\beta})^2 - \frac{n}{2}\log(2\pi\sigma^2)\,.$$

- A Função Deviance fica

$$D = 2\left[l(\mathbf{b}_{max}, \mathbf{y}) - l(\mathbf{b}, \mathbf{y})\right]$$

ou seja

$$D = 2\left[-\frac{n}{2}\log(2\pi\sigma^2) - \left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mathbf{x}_i^T\hat{\beta})^2 - \frac{n}{2}\log(2\pi\sigma^2)\,.\right)\right]$$

**Exemplo: (continuação)**

- Ficamos então com

$$D = 2\left[\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mathbf{x}_i^T\hat{\beta})^2\right] = \frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i - \mathbf{x}_i^T\hat{\beta})^2$$

- Se desconsiderarmo so termo $1/\sigma^2$ esse termo corresponde a que medida em regressão linear?
- Soma de quadrados dos resíduos

$$e_i = y_i - \mathbf{x}_i^T\hat{\beta}\ .$$

## Exemplo:

- Considere uma amostra

$$Y_1, Y_2, \ldots, Y_n$$

independente tal que $Y_i \sim Poisson(\lambda_i)$.

- A função de verossimilhança é dada por

$$f(\mathbf{y}, \boldsymbol{\lambda}) = \prod_i \lambda_i^{y_i} \frac{e^{-\lambda_i}}{y_i!}$$

portanto a log-verossimilhança é dada por

$$\log(f(\mathbf{y}, \boldsymbol{\lambda})) = \sum_i (y_i \log(\lambda_i) - \lambda_i - \log(y_i!))$$

$$\sum_i y_i \log(\lambda_i) - \sum_i \lambda_i - \sum_i \log(y_i!) .$$

- Vamos econtrar a Deviance para esse modelo.

**Exemplo: (continuação)**

- Para o modelo saturado temos que a média estimada pelo modelo é exatamente igual a observação.
- Portanto

$$\hat{E}(Y_i) = \hat{\lambda}_i = y_i \, .$$

- Vamos substituir $\lambda_i$ por $y_i$ na log-verossimilhança

$$\log(f(\mathbf{y};\beta)) = \sum_i y_i \log(y_i) - \sum_i y_i - \sum_i \log(y_i!) \, .$$

## Exemplo: (continuação)

- Vamos denotar por $\hat{\lambda}_i$ a média estimada pelo modelo sob análise.
- Para obtermos a log-verossimilhança no modelo sob análise basta substituir esse valor na equação

$$\sum_i y_i \log(\hat{\lambda}_i) - \sum_i \hat{\lambda}_i - \sum_i \log(y_i!) \, .$$

- Temos então que

$$l(\mathbf{b}_{max}, \mathbf{y}) = \sum_i y_i \log(y_i) - \sum_i y_i - \sum_i \log(y_i!)$$

$$l(\mathbf{b}, \mathbf{y}) = \sum_i y_i \log(\hat{\lambda}_i) - \sum_i \hat{\lambda}_i - \sum_i \log(y_i!)$$

**Exemplo: (continuação)**

- A função Deviance fica

$$D = 2\left[l(\mathbf{b}_{max}, \mathbf{y}) - l(\mathbf{b}, \mathbf{y})\right]$$

$$= \sum_i y_i \log(y_i) - \sum_i y_i - \sum_i \log(y_i!) -$$

$$\left(\sum_i y_i \log(\hat{\lambda}_i) - \sum_i \hat{\lambda}_i - \sum_i \log(y_i!)\right)$$

$$= \sum_i y_i \log\left(\frac{y_i}{\hat{\lambda}_i}\right) + \sum_i (\hat{\lambda}_i - y_i)$$

**Exemplo: (continuação)**

- Vamos retomar um exemplo visto na aula anterior.
- A Tabela a seguir apresenta os dados observados

**Exemplo: (continuação)**

- A figura a seguir mostra o gráfico de disperão entre as duas variáveis.

**Exemplo:** **(continuação)**

- O modelo ajustado foi o seguinte

$$\hat{y}_i = 7,4516 + 4,9353x_i$$

- A tabela a seguir mostra os valores necessários para o cálculo da Deviance

**Exemplo: (continuação)**

- A Deviance é dada por

$$D = 2\,[0,94735 - 0] = 1,8947\,.$$

- Você acha que o modelo está bem ajustado?
- Aparentemente sim, pois a Deviance é pequena.

- A Deviance é sempre maior ou igual a zero.
- Quando aumentamos o número de variáveis o que acontece?
- O ajuste do modelo aos dados melhora.
- A **Deviance diminui**.
- Qual a Deviance do modelo saturado? Zero.
- Então esse é o modelo ideal? Não.
- O modelo tem muitos parâmetros.

- Precisamos verificar se:
  - a redução na Deviance justifica inclusão de uma nova variável.
- Vejamos agora como testar adequação do modelo usando a Deviance.
- Pode-se mostrar que, se o modelo é adequado

$$D \sim \chi^2_{n-p}$$

onde
  - $n$ é o tamanho da amostra;
  - $p$ é o número de parâmetros do modelo.

## Exemplo:

- Considere o exemplo da Regressão de Poisson.
- O valor calculado da Deviance foi de

$$D = 1,8947\ .$$

- Nesse caso $p = 2$ e $n = 9$.
- Se o modelo está bem ajustado

$$D \sim \chi^2_7\ .$$

- Qual o valor esperado da distribuição? 7.
- Você acha que o modelo está bem ajustado?
- O modelo não é bem ajustado para valores pequenos ou grandes de $D$? Grandes.
- A Deviance observada é bem menor que o esperado, o modelo está bem ajustado.
- A Deviance é pequena, se comparada com o que seria esperado. $\Rightarrow$ Modelo bem ajustado!

## Exemplo:

- Hipóteses:

$$H_0 : \text{O modelo está bem ajustado (Deviance Pequena)}$$

$$H_1 : \text{O modelo está mal ajustado (Deviance Grande)}$$

- O teste é da forma:
  - se $D < \chi^2_C$ o modelo está bem ajustado;
  - se $D > \chi^2_C$ o modelo está mal ajustado (Rejeita $H_0$).
- Rejeita para valores grandes da Deviance.
- Teste unilateral à direita.
- Considerando $\alpha = 5\%$, o valor crítico é

$$\chi^2_C = 0,014 \quad \text{pois} P(\chi^2_7 < 0,014) = 0,95\,.$$

**Exemplo:**

- A figura a seguir mostra a densidade da $\chi^2_7$ com os valores observados e crítico em destaque.
- Rejeita-se $H_0$.

## Exemplo:

- O código em R usado para fazer o gráfico encontra-se a seguir.

```
curve(dchisq(x,7),from=0,to=20,ylab=expression(t
,xlab=expression(f(theta)))
abline(v=1.8947,col='red')
abline(v=0.014,col='blue')
legend("topright",c("Valor observado","Valor crí
, fill=c("red","blue"))
```

## Observação

- Além de testarmos o ajuste do modelo, podemos testar a significância dos parâmetros individualmente.
- Isso é feito usando a distribuição assintótica

$$\mathbf{b} \sim N(\beta, I(\beta^{-1}))$$

- Veremos como fazer esse teste no R na próxima aula..