



GET00211 - Modelos Lineares 2

Rafael Erbisti

Instituto de Matemática e Estatística
Universidade Federal Fluminense

Aula 06

Adequabilidade de um modelo

- A avaliação da adequação de um modelo pode ser realizada por meio da comparação com um modelo mais geral, caracterizado por conter o número máximo de parâmetros que podem ser estimados.
- A comparação pode ser feita através das verossimilhanças baseadas no modelo de interesse com aquela do **modelo maximal**, **saturado** ou **completo**.
- Este modelo é definido como:
 - um MLG que possui a mesma distribuição que a do modelo de interesse;
 - tem a mesma função de ligação que o modelo de interesse;
 - o número de parâmetros no modelo completo é igual ao total de observações n .



Adequabilidade de um modelo

- Com n observações disponíveis, é possível ajustar modelos que incluam no máximo n parâmetros.
- O modelo mais simples, denominado **modelo nulo**, possui apenas um parâmetro, que representa uma média comum para os Y 's.
- No outro extremo, o **modelo completo** possui n parâmetros, um para cada observação, de modo que os μ 's estimados reproduzem exatamente os dados observados.
- Algumas observações podem compartilhar o mesmo componente linear ou padrão de covariáveis, isto é, correspondem à mesma combinação de níveis de fatores e aos mesmos valores de variáveis explicativas contínuas.
- Essas observações são chamadas de **réplicas**.
- Nesses casos, o número máximo de parâmetros que podem ser estimados no modelo saturado é igual ao número de combinações lineares potencialmente distintas, o que pode ser menor que n .



Adequabilidade de um modelo

- Na prática, o **modelo nulo** é muito simples, enquanto o **modelo completo** é não informativo, pois apenas reproduz os dados sem resumi-los.
- Entretanto, o modelo completo serve como base para medir a discrepância de modelos intermediários, com p parâmetros.

Exemplo: Suponha $Y_i \sim \text{Poisson}(\mu_i)$, para $i = 1, \dots, n$.

- **Modelo nulo:** $\log(\mu_i) = \beta_0$, para todo i (apenas uma taxa média comum).
- **Modelo intermediário:** $\log(\mu_i) = \beta_0 + \beta_1 x_i$ (regressão de Poisson com $p = 2$).
- **Modelo completo:** $\log(\mu_i) = \beta_i$, com um parâmetro diferente para cada observação ($p = n$).



Comparação de verossimilhanças

- Sejam $\hat{\beta}_{\max}$ e $\hat{\beta}$ os EMVs do modelo completo e do modelo de interesse, respectivamente.
- Podemos calcular $L(\hat{\beta}_{\max}; \mathbf{y})$ e $L(\hat{\beta}; \mathbf{y})$.
- A função de verossimilhança do modelo saturado avaliada em $\hat{\beta}_{\max}$, $L(\hat{\beta}_{\max}; \mathbf{y})$, será maior do que qualquer outra função de verossimilhança para essas observações (mesma distribuição e função de ligação), pois fornece a descrição mais completa dos dados.
- Se o modelo de interesse descreve bem os dados, então $L(\hat{\beta}; \mathbf{y})$ deve estar próximo de $L(\hat{\beta}_{\max}; \mathbf{y})$.
- Se o modelo não fornece um bom ajuste, então $L(\hat{\beta}; \mathbf{y})$ será significativamente menor que $L(\hat{\beta}_{\max}; \mathbf{y})$.



Exemplo: Modelo Poisson

- Modelo de interesse:

$$\log(\mu_i) = \beta_0 + \beta_1 x_i$$

Estimamos $\hat{\beta}_0$ e $\hat{\beta}_1$ por máxima verossimilhança (Método Escore), ou seja, $\hat{\mu}_i = \exp\{\hat{\beta}_0 + \hat{\beta}_1 x_i\}$.

- Modelo saturado (completo):

$$\log(\mu_i) = \beta_i$$

Estimamos $\hat{\beta}_i = \log(y_i)$, ou seja, cada $\hat{\mu}_i = y_i$.

Função de verossimilhança da Poisson:

$$L(\boldsymbol{\mu}; \mathbf{y}) = \sum_{i=1}^n [y_i \log(\mu_i) - \mu_i - \log(y_i!)]$$

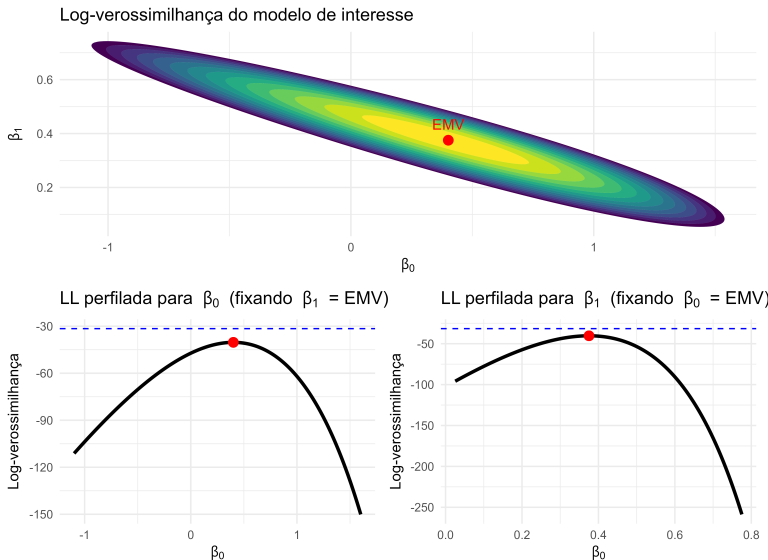
- No modelo saturado, a verossimilhança é máxima:

$$\hat{\mu}_i = y_i \implies L(\hat{\boldsymbol{\mu}}; \mathbf{y}) \text{ é máxima}$$

- A verossimilhança do modelo de interesse deve ser próxima à do modelo saturado quando o ajuste é adequado.



Exemplo: Modelo Poisson



Adequabilidade de um modelo

- Faz sentido, então, utilizar alguma medida que compare a verossimilhança do modelo de interesse com a verossimilhança do modelo saturado.

- Utiliza-se a estatística

$$\lambda = \frac{L(\hat{\beta}_{\max}; \mathbf{y})}{L(\hat{\beta}; \mathbf{y})}$$

como uma medida de bondade de ajuste.

- Equivalentemente, pode-se usar

$$\log \lambda = \ell(\hat{\beta}_{\max}; \mathbf{y}) - \ell(\hat{\beta}; \mathbf{y}).$$

- Valores altos de $\log \lambda$ sugerem que o modelo de interesse fornece uma descrição pobre dos dados relativa ao modelo completo.



Distribuição amostral para a estatística log-verossimilhança

- Suponha que o modelo de interesse envolve p parâmetros, descritos no vetor β .
- Considere uma aproximação para $\ell(\beta; \mathbf{y})$ através de uma expansão em série de Taylor, até o termo de segunda ordem, em torno do EMV $\hat{\beta}$:

$$\ell(\beta; \mathbf{y}) = \ell(\hat{\beta}; \mathbf{y}) + (\beta - \hat{\beta})' \mathbf{U}(\hat{\beta}) + \frac{1}{2}(\beta - \hat{\beta})' \mathbf{H}(\hat{\beta})(\beta - \hat{\beta})$$

- Como $\mathbf{U}(\hat{\beta}) = 0$ e usando a aproximação $\mathbf{H}(\hat{\beta}) \approx -\mathcal{I}$, temos:

$$\ell(\hat{\beta}; \mathbf{y}) - \ell(\beta; \mathbf{y}) \approx \frac{1}{2}(\hat{\beta} - \beta)' \mathcal{I}(\hat{\beta} - \beta)$$

- Mas já vimos que $(\hat{\beta} - \beta)' \mathcal{I}(\hat{\beta} - \beta) \sim \chi_p^2$, portanto:

$$2(\ell(\hat{\beta}; \mathbf{y}) - \ell(\beta; \mathbf{y})) \sim \chi_p^2$$



- A **deviance**, também chamada de estatística do log da razão de verossimilhanças, é definida como

$$D = 2 \log \lambda = 2(\ell(\hat{\beta}_{\max}; \mathbf{y}) - \ell(\hat{\beta}; \mathbf{y})).$$

- Somando e subtraindo $\ell(\beta_{\max}; \mathbf{y})$ e $\ell(\beta; \mathbf{y})$ na expressão acima, tem-se:

$$D = 2 \left\{ \left[\ell(\hat{\beta}_{\max}; \mathbf{y}) - \ell(\beta_{\max}; \mathbf{y}) \right] - \left[\ell(\hat{\beta}; \mathbf{y}) - \ell(\beta; \mathbf{y}) \right] + \left[\ell(\beta_{\max}; \mathbf{y}) - \ell(\beta; \mathbf{y}) \right] \right\}$$

- O primeiro termo tem distribuição χ_n^2 , onde n é o número de parâmetros do modelo completo.
- O segundo termo tem distribuição χ_p^2 , onde p é o número de parâmetros do modelo de interesse.
- O terceiro termo é uma constante positiva que será próxima de zero se o modelo com p parâmetros descrever os dados tão bem quanto o modelo completo.



- Se as variáveis definidas nos dois primeiros termos forem independentes e o terceiro termo for constante (próximo de zero), então

$$D \sim \chi_{n-p}^2$$

se o modelo for adequado.

- Se o modelo de interesse não for adequado, o terceiro termo será grande e, consequentemente,

$$D \sim \chi_{n-p}^2(\nu),$$

onde

$$\nu = \ell(\beta_{\max}; \mathbf{y}) - \ell(\beta; \mathbf{y})$$

é o parâmetro de não-centralidade da distribuição qui-quadrado não central.



Medindo a bondade de ajuste

- Se a variável resposta é normalmente distribuída, então D tem distribuição qui-quadrado exata.
- Contudo, nesse caso, D depende de $V(Y_i) = \sigma^2$, que usualmente é desconhecida.
- Isso significa que D não pode ser usada diretamente como estatística de bondade de ajuste.
- Para Y_i 's com outras distribuições, a distribuição amostral de D será aproximada.
- A distribuição amostral da estatística D pode ser usada para medir a adequabilidade de um modelo, comparando com a distribuição qui-quadrado apropriada.
- Se o modelo for bom, esperamos que D seja próximo do meio da distribuição.
- Isso é fácil de medir, já que a média de uma χ_m^2 é m .



Medindo a bondade de ajuste: Teste de hipóteses

- Podemos fazer testes de hipóteses sobre β utilizando a distribuição assintótica do estimador $\hat{\beta}$:

$$\hat{\beta} \sim N_p(\beta, \mathcal{I}^{-1})$$

ou, equivalentemente, usando a estatística de Wald:

$$(\hat{\beta} - \beta)' \mathcal{I}(\hat{\beta} - \beta) \sim \chi_p^2.$$

- Uma abordagem alternativa é **comparar a bondade de ajuste de dois modelos**, M_0 e M_1 .
- Os modelos precisam ser encaixados ou hierárquicos, isto é, eles têm a mesma distribuição de probabilidade e a mesma função de ligação, mas o preditor linear do modelo mais simples, M_0 , é um caso especial do preditor linear do modelo mais geral, M_1 .



Medindo a bondade de ajuste: Teste de hipóteses

- Considere a **hipótese nula** correspondendo ao modelo M_0 :

$$H_0 : \beta = \beta_0 = (\beta_1, \dots, \beta_q)'$$

e uma **hipótese alternativa** correspondendo ao modelo M_1 :

$$H_1 : \beta = \beta_1 = (\beta_1, \dots, \beta_p)', \quad \text{com } q < p < n.$$

- Podemos testar H_0 versus H_1 usando a estatística baseada na diferença das deviances:

$$\begin{aligned}\Delta D &= D_0 - D_1 \\ &= 2[\ell(\hat{\beta}_{\max}; \mathbf{y}) - \ell(\hat{\beta}_0; \mathbf{y}) - \ell(\hat{\beta}_{\max}; \mathbf{y}) + \ell(\hat{\beta}_1; \mathbf{y})] \\ &= 2[\ell(\hat{\beta}_1; \mathbf{y}) - \ell(\hat{\beta}_0; \mathbf{y})].\end{aligned}$$

- Se ambos os modelos descrevem bem os dados, então

$$D_0 \sim \chi_{n-q}^2 \quad \text{e} \quad D_1 \sim \chi_{n-p}^2,$$

de modo que, dado que certas condições de independência sejam válidas,

$$\Delta D \sim \chi_{p-q}^2.$$



Medindo a bondade de ajuste: Teste de hipóteses

- Se o valor de ΔD for consistente com a distribuição χ^2_{p-q} , escolheríamos o modelo correspondente a H_0 por ser o mais simples.
- Se o valor de ΔD estiver na região crítica (isto é, maior do que o limite superior do $100\alpha\%$ da distribuição χ^2_{p-q}), rejeitaríamos H_0 em favor de H_1 , baseados no fato de que β_1 fornece um ajuste significativamente melhor na descrição dos dados.
- Em geral, a distribuição amostral de ΔD é melhor aproximada pela distribuição qui-quadrado do que a distribuição da estatística individual D .



Estatística Qui-Quadrado de Pearson

- A estatística de **Pearson** é usada para avaliar a bondade de ajuste de modelos, especialmente em tabelas de contingência ou modelos de regressão para dados categóricos.

- Definição:

$$\chi^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i}$$

onde:

- O_i = valor observado da i -ésima categoria
 - E_i = valor esperado da i -ésima categoria segundo o modelo
 - n = número total de observações ou células
-
- Sob a hipótese de que o modelo se ajusta bem, χ^2 segue aproximadamente uma distribuição χ^2 com $n - p$ graus de liberdade, onde p é o número de parâmetros estimados.
-
- Valores altos de χ^2 indicam que o modelo não se ajusta bem aos dados.



Comparação entre Deviance e Qui-Quadrado de Pearson

- A escolha entre D e X^2 depende da adequabilidade da aproximação à distribuição χ^2_{n-p} .
- Há evidências que sugerem que, em geral, X^2 é mais robusta do que D , uma vez que D é influenciada por valores pequenos das frequências.
- Todas as aproximações tendem a ser pobres se as frequências esperadas forem muito pequenas.



Outros critérios para verificar bondade de ajuste

- Em analogia ao R^2 para regressão Normal, outra estatística utilizada é o **pseudo- R^2** :

$$\text{pseudo-}R^2 = \frac{\ell(\tilde{\beta}; \mathbf{y}) - \ell(\hat{\beta}; \mathbf{y})}{\ell(\tilde{\beta}; \mathbf{y})} = 1 - \frac{\ell(\hat{\beta}; \mathbf{y})}{\ell(\tilde{\beta}; \mathbf{y})}$$

que representa ganhos proporcionais na função log-verossimilhança devido aos termos do modelo de interesse ($\ell(\hat{\beta}; \mathbf{y})$), quando comparado ao modelo nulo ($\ell(\tilde{\beta}; \mathbf{y})$).

- É preferível o modelo com maior pseudo- R^2 .



Outros critérios para verificar bondade de ajuste

- O Critério de Informação de Akaike (AIC) e o Critério de Informação Bayesiano (BIC) são outras estatísticas de qualidade de ajuste baseadas na função verossimilhança, com penalização para o número de parâmetros estimados e para a quantidade de dados.

- São definidos como:

$$\text{AIC} = -2 \ell(\hat{\beta}; \mathbf{y}) + 2p$$

$$\text{BIC} = -2 \ell(\hat{\beta}; \mathbf{y}) + p \log(n)$$

onde p é o número de parâmetros estimados e n é o total de observações.

- Modelos com menores AIC e BIC são preferíveis.

