Seleção e comparação de modelos na classe dos MLG

Prof. Caio Azevedo

Seleção de modelos

- Vimos como verificar se um determinado modelo se ajusta adequadamente aos dados.
- Uma outra questão de interesse surge quando se dispõe de diversos modelos (que se ajustam adequadamente aos dados) e respondem às perguntas de interesse, e queremos escolher um como o “mais apropriado”.
- Há diversas técnicas disponíveis para este fim.
- Veremos técnicas baseadas em testes de hipótese e comparação de estatísticas de qualidade de ajuste.

Teste da razão de verossimilhanças

- Sejam M_1 e M_2 dois modelos, em que M_1 está encaixado em M_2, ou seja, o modelo M_1 é um caso particular de M_2.
- Por exemplo, M_1 é um modelo linear obtido de M_2, o qual é um modelo quadrático.
- Neste caso temos que

  H_0 : **o modelo** M_1 **é preferível ao modelo** M_2 **vs** H_1 : **o modelo** M_2 **é preferível ao modelo** M_1.

Teste da razão de verossimilhanças (cont.)

- Seja $\widehat{\boldsymbol{\theta}}_i$ o estimador de máxima verossimilhança obtido sob o modelo i e $\widetilde{\boldsymbol{\theta}}_i$ sua respectiva estimativa.
- Denote por $L_i(\widehat{\boldsymbol{\theta}})$ e $l_i(\widehat{\boldsymbol{\theta}})$ o máximo da verossimilhança e da log-verossimilhança do modelo i, respectivamente, em relação aos estimadores enquanto que $L_i(\widetilde{\boldsymbol{\theta}})$ e $l_i(\widetilde{\boldsymbol{\theta}})$ são os respectivos máximos avaliados nas estimativas.

Teste da razão de verossimilhanças (cont.)

- A estatística do TRV é dada por $\Delta = \frac{L_1(\widehat{\boldsymbol{\theta}}_1)}{L_2(\widehat{\boldsymbol{\theta}}_2)}$.
- Rejeita-se H_0 se $\Delta \geq \delta_c$, em que δ_c é um valor crítico adequado.
- Alternativamente, rejeitamos H_0 se

$$\Lambda = -2ln(\Delta) = -2\left(l_1(\widehat{\boldsymbol{\theta}}_1) - l_2(\widehat{\boldsymbol{\theta}}_2)\right) \geq \lambda_c,$$

em que $P(Q \geq \lambda_c) = \alpha$, $Q \approx \chi^2_{(\gamma)}$ e

$\gamma =$ número de parâmetros do modelo M_2 - número de parâmetros do modelo M_1.

- Nesse caso, $p - valor \approx P(Q \geq \lambda | H_0)$, em que λ é o valor observado da estatística Λ e $Q \sim \chi^2_{\gamma}$. Assim, rejeita-se H_0 se p-valor $\leq \alpha$.

Estatísticas de comparação de modelos

- O TRV é apropriado na comparação somente de modelos encaixados (o modelo com menor número de parâmetros é um caso particular do modelo com maior número de parâmetros).
- Além disso, ele não leva em consideração (diretamente) o número de parâmetros do modelo (somente na distribuição da estatística).
- Existem várias alternativas, em termos de estatísticas para comparar modelos, que "penalizam" a verossimilhança em relação ao número de parâmetros, tamanho da amostra entre outros fatores.
- Veremos o AIC e o BIC.

Estatísticas de comparação de modelos (cont.)

- O AIC e BIC, para o i-ésimo modelo, são dados, respectivamente, por:

$$
\begin{aligned}
AIC_i &= -2l_i(\widetilde{\theta}_i) + 2k \\
BIC_i &= -2l_i(\widetilde{\theta}_i) + k\ln(n)
\end{aligned}
$$

  que $l_i(\widetilde{\theta}_i)$ denota a log-verossimilhança do i-ésimo modelo avaliada em alguma estimativa (p.e. máxima verossimilhança), k é o número de parâmetros e n é o número de observações.

- Portanto, o modelo que apresentar os menores valores, será o modelo "melhor ajustado" aos dados.

Métodos de seleção "dinâmico" ou automatizados

- Existem métodos que selecionam modelos, fixados alguns critérios, de modo "dinâmico" (automatizado).
- Veremos os métodos "forward", "backward" e "stepwise".
- Tais métodos são particularmente úteis quanto se dispões de muitas covariáveis e/ou muitos fatores.
- Sem perda de generalidade, vamos considerar um determinado modelo (normal linear, linear generalizado) tal que o preditor linear é dado por

$$\eta_{ij} = \beta_0 + \sum_{j=1}^{p-1} \beta_j x_{ij}$$

Método "forward"

- Primeiramente, ajustamos um modelo com somente o intercepto, ou seja $\eta_{ij} = \beta_0$. Ajustamos então, para cada variável explicativa, um modelo

$$\eta_{ij} = \beta_0 + \beta_j x_{ij}, j = 1, 2, ..., p - 1$$

- Testa-se $H_0 : \beta_j = 0$ vs $H_1 : \beta_j \neq 0$, j=1,2,...,p-1 (usando-se algum teste como o TRV, teste $\mathbf{C}\boldsymbol{\beta}$, ou alguma estatística de comparação de modelos). Seja P o menor nível descritivo entre os $p - 1$ testes. Se $P \leq P_E$ a variável correspondente entra no modelo (caso contrário, o processo é interrompido).

Métodos “forward” (cont.)

- Vamor supor que a variável X_1 foi escolhida. Então, no passo seguinte, ajustamos os modelos

$$\eta_{ij} = \beta_0 + \beta_1 x_{1j} + \beta_j x_{ij}, j = 2, ..., p-1$$

- Testa-se $H_0 : \beta_j = 0$ vs $H_1 : \beta_j \neq 0$, j=2,...,p-1 (usando-se algum teste como TRV, teste $\mathbf{C}\boldsymbol{\beta}$, ou alguma estatística de comparação de modelos). Seja P o menor nível descritivo entre os $p-2$ testes. Se $P \leq P_E$ a variável correspondente entra no modelo. Repetimos o procedimento até que ocorra $P > P_E$.

Método "backward"

- Primeiramente, ajustamos o seguinte modelo:

$$\eta_{ij} = \beta_0 + \sum_{j=1}^{p-1} \beta_j x_{ij}$$

- Testa-se $H_0 : \beta_j = 0$ vs $H_1 : \beta_j \neq 0$, j=1,2,...,p-1 (usando-se algum teste como o TRV, teste $\mathbf{C}\boldsymbol{\beta}$, ou alguma estatística de comparação de modelos). Seja P o maior nível descritivo entre os $p-1$ testes. Se $P > P_S$ a variável correspondente sai do modelo (caso contrário, o processo é interrompido).

Método “backward” (cont.)

- Vamos supor que X_1 tenha saído do modelo. Então ajustamos o seguinte modelo

$$\eta_{ij} = \beta_0 + \sum_{j=2}^{p-1} \beta_j x_{ij}$$

- Testa-se $H_0 : \beta_j = 0$ vs $H_1 : \beta_j \neq 0$, j=2,...,p-1 (usando-se algum teste como TRV, teste $\mathbf{C}\boldsymbol{\beta}$, ou alguma estatística de comparação de modelos). Seja P o maior nível descritivo entre os $p-2$ testes. Se $P > P_S$ a variável correspondente sai do modelo. Repetimos o procedimento até que ocorra $P \leq P_S$.

Método "stepwise"

- É uma mistura dos dois procedimentos anteriores.
- Iniciamos o processo com o modelo $\eta_{ij} = \beta_0$. Após duas variáveis terem sido incluídas no modelo, verificamos se a primeira sai ou não do modelo.
- O processo continua até que nenhuma variável seja incluída ou retirada do modelo.
- Geralmente adotamos $0,15 \leq P_E, P_S \leq 0,25$. Outra possibilidade é usar $P_E = P_S = 0,20$.

Métodos anteriores usando AIC/BIC

- Para qualquer um dos métodos anteriores, se usarmos alguma estatística de comparação de modelos (como AIC ou BIC), procedemos da seguinte forma
  - Sempre escolhemos o modelo (retirar/incluir a variável) que apresentar o menor valor da estatística.
  - O processo é interrompido quando as estatísticas para todos os modelos possíveis aumentarem em relação ao modelo corrente.
- Observação: as estatísticas AIC e BIC também servem para comparar modelos que difiram em termos da função de ligação e distribuição da variável resposta, entre outras características.