

Estimação e seleção estrutural em modelos hierárquicos

Prof. Caio Azevedo

Introdução

- Recordem-se dos exemplos vistos anteriormente: [aqui](#) e [aqui](#).
- Nos modelos hierárquicos (MH) de dois níveis, podemos pensar em dois tipos de tamanhos de amostra:
 - O número de UAE's no nível 1 ($n_j, j = 1, 2, \dots, J$) (e.g., alunos, medidas repetidas, número de sub-parcelas etc), dentro de cada UAE do nível 2.
 - O número de UAE's no nível 2 (J) (eg, escolas, indivíduos, número de parcelas etc).

Introdução

- Se $n_j \rightarrow \infty$, para pelo menos um j , tem-se mais “informação” (veja também [link 1](#), [link 2](#)) para a mesma quantidade de parâmetros a serem estimados.
- Contudo, se $J \rightarrow \infty$, teremos mais efeitos aleatórios, ou seja, mais “parâmetros” para estimar, ainda que mais “informação” se torne disponível.
- Nos MH, usualmente, os efeitos fixos são chamados de parâmetros estruturais (não aumentam com o aumento do tamanho da amostra), enquanto que os efeitos aleatórios são chamados de parâmetros incidentais (podem aumentar com o aumento do tamanho da amostra).

Introdução

- Quando há somente parâmetros estruturais, sob certas Condições de Regularidade (CR), os estimadores de MV são consistentes e assintoticamente normais ([link](#)).
- A presença de parâmetros **incidentais** (veja também [link](#)) tende a violar as CR. Como tratar essa questão?
- **Neyman & Scott (1948)** discutem a consistência dos estimadores de máxima verossimilhança (EMV) dos parâmetros estruturais, na presença de parâmetros incidentais.

Introdução

- Algumas abordagens possíveis:
 - Considerar os estimadores de máxima verossimilhança marginal (integrando a verossimilhança conjunta com relação aos efeitos aleatórios) ([link](#)), dos efeitos fixos (parâmetros de regressão e componentes da variância) e depois prever os efeitos aleatórios utilizando procedimentos Bayesianos empíricos ([link](#)) - Máxima Verossimilhança Marginal (MVM).
 - Considerar os efeitos aleatórios como **dados faltantes** e utilizar algum tipo de **algoritmo EM** - Máxima verossimilhança via algoritmo EM.
 - Utilizar métodos Bayesianos marginais ([link](#)), os quais são uma versão Bayesiana da MVM ou plenos ([link](#)) (trabalhando com a posteriori conjunta) - Inferência Bayesiana.

Máxima Verossimilhança Marginal

- Em grande parte do curso, utilizaremos uma versão aprimorada do método de MVM, conhecida como máxima verossimilhança restrita ([MVR](#)).
- Uma forma de utilizar essa abordagem consiste em escrever o MH como um MMI (e trabalhar com uma modificação da verossimilhança marginal), como discutiu-se na aula de [modelos hierárquicos de dois níveis](#).
- Detalhes sobre propriedades desses estimadores são discutidas em [Demidenko 1](#), [Demidenko 2](#) e [Jiang](#), por exemplo

Modelo com dois níveis: formulação matricial geral

- Essencialmente, qualquer modelo hierárquico de dois níveis pode ser escrito, matricialmente, da seguinte forma (em que $j = 1, \dots, J$ definem os grupos - nível 2 da hierarquia).

$$\mathbf{Y}_{j(n_j \times 1)} = \mathbf{X}_{j(n_j \times p)} \boldsymbol{\beta}_{j(p \times 1)} + \boldsymbol{\xi}_{j(n_j \times 1)} \quad (1)$$

$$\boldsymbol{\beta}_{j(p \times 1)} = \mathbf{W}_{j(p \times q)} \boldsymbol{\gamma}_{(q \times 1)} + \mathbf{u}_{j(p \times 1)}, \quad (2)$$

em que \mathbf{Y}_j é o vetor de respostas do j ésimo conglomerado, \mathbf{X}_j e \mathbf{W}_j são matrizes de planejamento (conhecidas e não aleatórias), $\boldsymbol{\xi}_j \stackrel{ind.}{\sim} N_{n_j}(0, \boldsymbol{\Sigma}_j)$, $\mathbf{u}_j \stackrel{ind.}{\sim} N_p(0, \boldsymbol{\Psi})$, $\boldsymbol{\xi}_j \perp \mathbf{u}_j, \forall j$ e $\boldsymbol{\gamma}$ é um vector de efeitos fixos.

Relação entre modelos mistos (MMI) e modelos hierárquicos (MH)

- Podemos escrever a formulação (1) - (2) da seguinte forma:

$$\begin{aligned} \mathbf{Y}_j &= \mathbf{X}_j(\mathbf{W}_j\boldsymbol{\gamma} + \mathbf{u}_j) + \boldsymbol{\xi}_j \\ &= \mathbf{X}_j\mathbf{W}_j\boldsymbol{\gamma} + \mathbf{X}_j\mathbf{u}_j + \boldsymbol{\xi}_j \\ &= \mathbf{W}_j^*\boldsymbol{\gamma} + \mathbf{X}_j\mathbf{u}_j + \boldsymbol{\xi}_j = \mathbf{Z}_j\boldsymbol{\gamma} + \mathbf{X}_j\mathbf{u}_j + \boldsymbol{\xi}_j \end{aligned} \quad (3)$$

em que $\mathbf{W}_j^* = \mathbf{Z}_j = \mathbf{X}_j\mathbf{W}_j$. A Equação (3) **assemelha-se** à estrutura (matricial) de modelos mistos veja ([link](#)). Contudo, existem diferenças conceituais e metodológicas entre eles ([como já discutido](#)).

Formas matriciais gerais

Modelo Hierárquico de dois níveis

$$\mathbf{Y}_{(n \times 1)} = \mathbf{X}_{(n \times Jp)} \boldsymbol{\beta}_{(Jp \times 1)} + \boldsymbol{\xi}_{(n \times 1)}$$

$$\boldsymbol{\beta}_{(Jp \times 1)} = \mathbf{W}_{(Jp \times q)} \boldsymbol{\gamma}_{(q \times 1)} + \mathbf{u}_{(Jp \times 1)}$$

(Respectivo) Modelo Misto

$$\mathbf{Y}_{(n \times 1)} = \mathbf{Z}_{(n \times q)} \boldsymbol{\gamma}_{(q \times 1)} + \mathbf{X}_{(n \times Jp)} \mathbf{u}_{(Jp \times 1)} + \boldsymbol{\xi}_{(n \times 1)}$$

$$\mathbf{Y} = (\mathbf{Y}'_1, \mathbf{Y}'_2, \dots, \mathbf{Y}'_J)', \quad \mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_J), \quad \boldsymbol{\beta} = (\boldsymbol{\beta}'_1, \boldsymbol{\beta}'_2, \dots, \boldsymbol{\beta}'_J)',$$

$\boldsymbol{\xi} = (\boldsymbol{\xi}'_1, \boldsymbol{\xi}'_2, \dots, \boldsymbol{\xi}'_J)'$ $\mathbf{u} = (\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_J)'$ (em que todos esses vetores e matrizes são formadas pela concatenação vertical dos vetores correspondentes)

Algumas propriedades do modelo

- $\mathcal{E}(\mathbf{Y}_j|\mathbf{u}_j) = \mathbf{Z}_j\boldsymbol{\gamma} + \mathbf{X}_j\mathbf{u}_j.$
- $\mathcal{E}(\mathbf{Y}_j) = \mathbf{Z}_j\boldsymbol{\gamma}.$
- $\text{Cov}(\mathbf{Y}_j|\mathbf{u}_j) = \boldsymbol{\Sigma}_j.$
- $\text{Cov}(\mathbf{Y}_j) = \mathbf{V}_j = \mathbf{X}_j\boldsymbol{\Psi}\mathbf{X}_j' + \boldsymbol{\Sigma}_j.$
- $\mathbf{Y}_j|\mathbf{u}_j \sim N_{n_j}(\mathbf{Z}_j\boldsymbol{\gamma} + \mathbf{X}_j\mathbf{u}_j, \boldsymbol{\Sigma}_j).$ Além disso, como

$$\begin{aligned}\mathbf{Y}_j|\mathbf{u}_j &\stackrel{ind.}{\sim} N_{n_j}(\mathbf{Z}_j\boldsymbol{\gamma} + \mathbf{X}_j\mathbf{u}_j, \boldsymbol{\Sigma}_j) \\ \mathbf{u}_j &\stackrel{iid}{\sim} N_q(0, \boldsymbol{\Psi})\end{aligned}$$

portanto

$$\mathbf{Y}_j \stackrel{ind.}{\sim} N_{n_j}(\mathbf{Z}_j\boldsymbol{\gamma}, \mathbf{X}_j\boldsymbol{\Psi}\mathbf{X}_j' + \boldsymbol{\Sigma}_j)$$

Estruturas para as matrizes de covariância

- Diferentes escolhas para Ψ e Σ_j induzem diferentes estruturas de dependência para o vetor de respostas.
- Por exemplo, quando $\Sigma_j = \sigma^2 I_{n_j}$, tem-se o modelo de independência condicional homocedástico. Modelos de independência condicional são bastante considerados em psicometria (Teoria de Resposta ao item).
- Por outro lado, quando $\Sigma_j = \sigma^2 I_{n_j}$ e $\Psi \equiv 0$, tem-se o modelo de regressão linear usual (homocedástico e com as observações independentes).

Estruturas para as matrizes de covariância

- Dependendo da importância (interpretação) dos efeitos aleatórios para o estudo, podemos pensar em diferentes estruturas de covariância para eles.
- Existem diversas técnicas para sugestão/escolha de matrizes de covariâncias, veja [Rocha](#) e [Dunson](#), por exemplo.
- Parcimônia deve ser sempre considerada, nesse processo.

Modelos para a estrutura de covariância

- Usualmente, escolheremos para Ψ uma matriz não estruturada (sem um padrão específico, ou seja, variâncias e correlações livres, como [aqui](#)) e $\Sigma_j = \sigma^2 I_{n_j}$.
- A matriz de variâncias e covariâncias do vetor de respostas será, portanto, uma combinação das matrizes escolhidas para os erros e para os efeitos aleatórios.
- Naturalmente, as escolhas são limitadas pelos recursos computacionais a serem utilizados.
- Diversas estruturas podem ser encontradas, por exemplo, no manual do [PROC MIXED SAS](#) e [aqui](#).

Estimação

- Sob a ótica frequentista, em geral, trabalha-se com a distribuição marginal de \mathbf{Y}_j em relação à \mathbf{u}_j , ou seja $\mathbf{Y}_j \sim N_{n_j}(\mathbf{Z}_j\boldsymbol{\gamma}, \mathbf{X}_j\boldsymbol{\Psi}\mathbf{X}_j' + \boldsymbol{\Sigma}_j)$
- Alternativa: **algoritmo EM** utilizando a distribuição conjunta de $(\mathbf{Y}, \mathbf{u})'$.
- Também existem **métodos Bayesianos**.
- Suposição : $\boldsymbol{\Sigma}_j = g(\boldsymbol{\theta}_1)$ e $\boldsymbol{\Psi} = h(\boldsymbol{\theta}_2)$ de modo que $\boldsymbol{\theta}_1$ e $\boldsymbol{\theta}_2$ não possuem componentes comuns, em que $\boldsymbol{\theta}_1$ e $\boldsymbol{\theta}_2$ tem r_1 e r_2 parâmetros, respectivamente.

Cont.

- Log-verossimilhança (marginal) para J observações:

$$l(\boldsymbol{\gamma}, \boldsymbol{\theta}) = -\frac{1}{2} \ln(2\pi) \sum_{j=1}^J n_j - \frac{1}{2} \sum_{j=1}^J \ln |\mathbf{V}_j| + -\frac{1}{2} \sum_{j=1}^J (\mathbf{Y}_j - \mathbf{Z}_j \boldsymbol{\gamma})' \mathbf{V}_j^{-1} (\mathbf{Y}_j - \mathbf{Z}_j \boldsymbol{\gamma}) \quad (4)$$

$$\mathbf{V}_j \equiv \mathbf{V}_j(\boldsymbol{\theta}) = \mathbf{X}_j \boldsymbol{\Psi}(\boldsymbol{\theta}_2) \mathbf{X}_j' + \boldsymbol{\Sigma}_j(\boldsymbol{\theta}_1), \quad \boldsymbol{\theta} = (\boldsymbol{\theta}_1', \boldsymbol{\theta}_2')$$

- Se $\boldsymbol{\theta}_{(r \times 1)}$ ($r = r_1 + r_2$) for conhecido, o estimador de MVM (que corresponde ao estimador de MQG) de $\boldsymbol{\gamma}$ é dado por:

$$\hat{\boldsymbol{\gamma}} = \left(\sum_{j=1}^J \mathbf{Z}_j' \mathbf{V}_j^{-1} \mathbf{Z}_j \right)^{-1} \left(\sum_{j=1}^J \mathbf{Z}_j' \mathbf{V}_j^{-1} \mathbf{Y}_j \right) \quad (5)$$

Cont.

- Para estimar (θ) , substituímos (5) em (4), obtendo uma **log-verossimilhança perfilada**:

$$l(\theta) = -\frac{1}{2} \ln(2\pi) \sum_{j=1}^J n_j - \frac{1}{2} \sum_{j=1}^J \ln |\mathbf{V}_j| - \frac{1}{2} \sum_{j=1}^J (\mathbf{Y}_j - \mathbf{Z}_j \hat{\gamma})' \mathbf{V}_j^{-1} (\mathbf{Y}_j - \mathbf{Z}_j \hat{\gamma}) \quad (6)$$

- A maximização da log-verossimilhança (6) tem de ser feita através de **métodos iterativos** como os algoritmos de Newton-Raphson, Escore de Fisher, Gauss-Newton, BFGS.
- Uma vez que tais estimativas forem obtidas, as inserimos em (5).

Cont.

- As distribuições assintóticas dos estimadores podem ser obtidas através de **TCL** (Teorema Central do Limite) apropriados.
- Os **erros-padrão assintóticos** podem ser obtidos através das inversas das informações de Fisher para θ e através de uma fórmula analítica (para γ).
- Os estimadores de MVM para γ são não viesados, mas o mesmo não acontece com os estimadores de MVM de θ (**Singer et al (2018)**).
- Alternativa: estimadores de MVM restritos (MVR) (também chamados de estimadores MV residuais).

Algoritmo (estimação por MVM)

- Primeiramente, estima-se θ através de algum algoritmo de maximização conveniente (NR, RF, Gauss-Newton, BFGS) para resolver o sistema de equações obtido através da derivação de (6) em relação a θ_k , ou seja (próximo slide):
- Os métodos implementados no R, por exemplo a função `optim` podem ser úteis (veja também [Howard](#) e [Métodos Computacionais em Estatística](#)).

Algoritmo (estimação por MVM)

$$\begin{aligned} S(\theta_k) &= \frac{\partial l(\boldsymbol{\theta})}{\partial \theta_k} = -\frac{1}{2} \sum_{j=1}^J \frac{\partial \ln |\mathbf{V}_j|}{\partial \theta_k} - \frac{1}{2} \sum_{j=1}^J \text{tr} \left[\frac{\partial \mathbf{V}_j^{-1}}{\partial \theta_k} (\mathbf{Y}_j - \mathbf{Z}_j \hat{\boldsymbol{\gamma}}) (\mathbf{Y}_j - \mathbf{Z}_j \hat{\boldsymbol{\gamma}})' \right. \\ &+ \left. \mathbf{V}_j^{-1} \frac{\partial (\mathbf{Y}_j - \mathbf{Z}_j \hat{\boldsymbol{\gamma}}) (\mathbf{Y}_j - \mathbf{Z}_j \hat{\boldsymbol{\gamma}})'}{\partial \theta_k} \right] = -\frac{1}{2} \sum_{j=1}^J \text{tr} \left[\mathbf{V}_j^{-1} \frac{\partial \mathbf{V}_j}{\partial \theta_k} \right] \\ &- \frac{1}{2} \sum_{j=1}^J \text{tr} \left[\frac{\partial \mathbf{V}_j^{-1}}{\partial \theta_k} (\mathbf{Y}_j - \mathbf{Z}_j \hat{\boldsymbol{\gamma}}) (\mathbf{Y}_j - \mathbf{Z}_j \hat{\boldsymbol{\gamma}})' \right. \\ &+ \left. \mathbf{V}_j^{-1} \frac{\partial (\mathbf{Y}_j - \mathbf{Z}_j \hat{\boldsymbol{\gamma}}) (\mathbf{Y}_j - \mathbf{Z}_j \hat{\boldsymbol{\gamma}})'}{\partial \theta_k} \right] \end{aligned}$$

Algoritmo (estimação por MVM)

- A notação $\frac{\partial \mathbf{V}_j}{\partial \boldsymbol{\theta}}$ representa a derivada de \mathbf{V}_j com relação à cada componente de $\boldsymbol{\theta}$ o que resulta, para cada componente, numa matriz.
- Com as estimativas de $\boldsymbol{\theta}$, digamos $\tilde{\boldsymbol{\theta}}$, obtem-se as estimativas de $\boldsymbol{\gamma}$, através de:

$$\tilde{\boldsymbol{\gamma}} = \left(\sum_{j=1}^J \mathbf{Z}_j' \mathbf{V}_j(\tilde{\boldsymbol{\gamma}})^{-1} \mathbf{Z}_j \right)^{-1} \left(\sum_{j=1}^J \mathbf{Z}_j' \mathbf{V}_j^{-1}(\tilde{\boldsymbol{\gamma}}) \mathbf{Y}_j \right)$$

- A matriz de covariâncias de $\hat{\boldsymbol{\gamma}}$ é dada por

$$\boldsymbol{\Sigma}_{\boldsymbol{\gamma}} = \left(\sum_{j=1}^J \mathbf{Z}_j' \mathbf{V}_j(\boldsymbol{\theta})^{-1} \mathbf{Z}_j \right)^{-1} \text{ e uma estimativa é dada por:}$$

$$\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{\gamma}} = \left(\sum_{j=1}^J \mathbf{Z}_j' \mathbf{V}_j(\tilde{\boldsymbol{\theta}})^{-1} \mathbf{Z}_j \right)^{-1}.$$

- Para o estimador $\hat{\theta}$ uma aproximação da matriz de covariâncias pode ser obtida através da inversa da matriz $-\mathbf{H}(\theta) = -\frac{\partial l(\theta)}{\partial \theta \partial \theta'}$ e uma estimativa é dada pela inversa de: $-\mathbf{H}(\tilde{\theta}) = -\frac{\partial l(\theta)}{\partial \theta \partial \theta'} \Big|_{\theta=\tilde{\theta}}$, respectivamente $\Sigma_{\theta} = -\mathbf{H}(\theta)^{-1}$ e $\tilde{\Sigma}_{\theta} = -\mathbf{H}(\tilde{\theta})^{-1}$.
- Os erros-padrão dos estimadores $\hat{\gamma}$ e $\hat{\theta}$ correspondem à raiz quadrada dos elementos da diagonal principal das respectivas matrizes de covariância.

- A distribuição dos estimadores (exata ou assintótica) pode ser obtida através de um dos seguintes métodos:
 - Convergência em distribuição dos estimadores de máxima verossimilhança ($\hat{\gamma} \approx N_q(\gamma, \Sigma_\gamma)$ e $\hat{\theta} \approx N_r(\theta, \Sigma_\theta)$), para $n_j, \forall j$ e n suficientemente grandes (assintótica).
 - Métodos de reamostragem (exata).
 - Método Delta (para funções, não lineares, dos parâmetros, que sejam de interesse) (assintótica).

Máxima verossimilhança restrita (ou residual)

- MVR: consiste em maximizar a verossimilhança de uma transformação ortogonal do vetor de respostas, ou seja, da verossimilhança induzida por $\mathbf{Y}^* = \mathbf{U}\mathbf{Y}$, em que $\mathbf{Y} = (\mathbf{Y}'_1, \dots, \mathbf{Y}'_J)'$.
- Em geral, escolhe-se $\mathbf{U} = \mathbf{I}_n - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$. Assim,

$\mathbf{Y}^* \sim N_{n-q}(0_n, \mathbf{U}\mathbf{V}\mathbf{U}')$, em que $\mathbf{V} = \text{Cov}(\mathbf{Y}) = \mathbf{X}\Psi\mathbf{X}' + \Sigma$,

$$\Sigma = \text{diag}(\Sigma_1, \Sigma_2, \dots, \Sigma_J) = \begin{bmatrix} \Sigma_1 & 0 & \dots & 0 \\ 0 & \Sigma_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \Sigma_J \end{bmatrix} \text{ e}$$

$$n = \sum_{j=1}^J n_j.$$

Máxima verossimilhança restrita (ou residual)

- A dimensão do vetor \mathbf{Y}^* ($n-q$) está relacionada ao fato de que $r(\mathbf{U}) = n - q$.
- Os estimadores de MVR de γ são não viesados enquanto que o viés do estimadores de MVR de θ são menores em comparação com os estimadores de MVM.
- O nome “residual” vem do fato de que a matriz \mathbf{U} gera alguns tipos de resíduos no ajuste por mínimos quadrados ordinários.

- A log-verossimilhança residual ou restrita é dada por

$$\begin{aligned}l_R(\boldsymbol{\theta}) &= -\frac{1}{2} \ln |\mathbf{V}| - \frac{1}{2} \ln |\mathbf{Z}' \mathbf{V}^{-1} \mathbf{Z}| \\ &\quad - \frac{n-q}{2} (\mathbf{Y} - \mathbf{Z}\hat{\boldsymbol{\gamma}})' \mathbf{V}^{-1} (\mathbf{Y} - \mathbf{Z}\hat{\boldsymbol{\gamma}}) - \frac{n-q}{2} \ln(2\pi).\end{aligned}$$

Cont.

- A log-verossimilhança residual ou restrita pode ser escrita como

$$\begin{aligned}l_R(\boldsymbol{\theta}) &= -\frac{1}{2} \ln(2\pi) \sum_{j=1}^J n_j - \frac{1}{2} \sum_{j=1}^J \ln |\mathbf{V}_j| \\ &- \frac{1}{2} \sum_{j=1}^J (\mathbf{Y}_j - \mathbf{Z}_j \hat{\boldsymbol{\gamma}})' \mathbf{V}_j^{-1} (\mathbf{Y}_j - \mathbf{Z}_j \hat{\boldsymbol{\gamma}}) \\ &- \frac{1}{2} \ln \left| \sum_{j=1}^n \mathbf{Z}_j' \mathbf{V}_j^{-1} \mathbf{Z}_j \right| + \text{const.}\end{aligned}\tag{7}$$

em que $\hat{\boldsymbol{\gamma}}$ é dado em (5).

Cont.

- Uma vez que os estimadores de MVR de θ forem obtidos, maximizando-se (7) (numericamente), os estimadores de MVR de γ podem ser obtidos inserindo aqueles em (5).
- As distribuições exatas ou assintóticas dos estimadores de MVR podem ser obtidas de modo semelhante aos dos estimadores de MVM.
- Lembrem-se de que estamos lidando com um conjunto de vetores aleatórios independentes mas não identicamente distribuídos $\mathbf{Y}_j \stackrel{ind.}{\sim} N_{n_j}(\mathbf{Z}_j\gamma, \mathbf{V}_j)$.
- TLC's que levem tal estrutura em consideração devem ser utilizados.

Algoritmo (estimação por MVR)

- Estima-se θ através de algum algoritmo de maximização conveniente (NR, RF, Gauss-Newton, BFGS), resolvendo-se o sistema de equações dado por:
- Os métodos implementados no R, por exemplo a função `optim` podem ser úteis (veja também [Howard](#) e [Métodos Computacionais em Estatística](#)).

Algoritmo (estimação por MVR)

$$\begin{aligned} S(\theta_k) &= \frac{\partial l_R(\boldsymbol{\theta})}{\partial \theta_k} = -\frac{1}{2} \sum_{j=1}^J \text{tr} \left[\mathbf{V}_j^{-1} \frac{\partial \mathbf{V}_j}{\partial \theta_k} \right] \\ &- \frac{1}{2} \sum_{j=1}^J \text{tr} \left[\frac{\partial \mathbf{V}_j^{-1}}{\partial \theta_k} (\mathbf{Y}_j - \mathbf{Z}_j \hat{\boldsymbol{\gamma}}) (\mathbf{Y}_j - \mathbf{Z}_j \hat{\boldsymbol{\gamma}})' \right. \\ &\left. + \mathbf{V}_j^{-1} \frac{\partial (\mathbf{Y}_j - \mathbf{Z}_j \hat{\boldsymbol{\gamma}}) (\mathbf{Y}_j - \mathbf{Z}_j \hat{\boldsymbol{\gamma}})'}{\partial \theta_k} \right] - \frac{1}{2} \sum_{i=1}^J \text{tr} \left[\mathbf{V}_j^{-1} \mathbf{Z}_j' \mathbf{V}_j^{-1} \frac{\partial \mathbf{V}_j^{-1}}{\partial \theta_k} \mathbf{V}_j^{-1} \mathbf{Z}_j \right] \end{aligned}$$

Algoritmo (estimação por MVR)

- Com as estimativas de θ , digamos $\tilde{\theta}_R$, obtem-se as estimativas de γ_R , ou seja:

$$\tilde{\gamma}_R = \left(\sum_{j=1}^J \mathbf{z}_j' \mathbf{V}_j(\tilde{\theta})^{-1} \mathbf{z}_j \right)^{-1} \left(\sum_{j=1}^J \mathbf{z}_j' \mathbf{V}_j^{-1}(\tilde{\theta}) \mathbf{y}_j \right)$$

- A matriz de covariâncias de $\hat{\gamma}_R$ é dada por

$$\Sigma_{\gamma_R} = \left(\sum_{j=1}^J \mathbf{z}_j' \mathbf{V}_j(\theta)^{-1} \mathbf{z}_j \right)^{-1} \text{ e uma estimativa é dada por:}$$

$$\tilde{\Sigma}_{\gamma_R} = \left(\sum_{j=1}^J \mathbf{z}_j' \mathbf{V}_j(\tilde{\theta})^{-1} \mathbf{z}_j \right)^{-1}.$$

- Para o estimador θ_R uma aproximação da matriz de covariâncias pode ser obtida através da inversa da matriz

$$\Sigma_{\theta_R} = -\mathbf{H}_R(\theta) = -\frac{\partial I_R(\theta)}{\partial \theta \partial \theta'} \text{ e uma estimativa é dada pela inversa de:}$$
$$-\mathbf{H}_R(\tilde{\theta}) = -\frac{\partial I_R(\theta)}{\partial \theta \partial \theta'} \Big|_{\theta=\tilde{\theta}}, \text{ respectivamente } \Sigma_{\theta_R} = -\mathbf{H}_R(\theta)^{-1} \text{ e}$$
$$\tilde{\Sigma}_{\theta_R} = -\mathbf{H}_R(\tilde{\theta})^{-1}.$$

- Os erros-padrão dos estimadores $\hat{\gamma}_R$ e $\hat{\theta}_R$ correspondem à raiz quadrada dos elementos da diagonal principal das respectivas matrizes de covariância.

- A distribuição dos estimadores pode ser obtida através de um dos seguintes métodos:
 - Convergência em distribuição dos estimadores de máxima verossimilhança (assintótica).
 - Métodos de reamostragem (exata).
 - Método Delta (assintótica).

Cont.

- Preditores para os efeitos aleatórios podem ser obtidos através da distribuição condicional (à posteriori), de $\mathbf{u}_j | \mathbf{y}_j$, ou seja

$$p(\mathbf{u}_j | \mathbf{y}_j) = \frac{p(\mathbf{y}_j | \mathbf{u}_j) p(\mathbf{u}_j)}{\int_{\mathbb{R}^p} p(\mathbf{y}_j | \mathbf{u}_j) p(\mathbf{u}_j) d\mathbf{u}}$$

a qual corresponde à (veja: [Lindley and Smith \(1972\)](#) e [curso de Infeência Bayesiana](#))

$$\mathbf{u}_j | \mathbf{y}_j \sim$$

$$N_p \left(\left(\mathbf{X}'_j \Sigma_j^{-1} \mathbf{X}_j + \Psi^{-1} \right)^{-1} \mathbf{X}'_j \Sigma_j^{-1} (\mathbf{y} - \mathbf{Z}\gamma), \left(\mathbf{X}'_j \Sigma_j^{-1} \mathbf{X}_j + \Psi^{-1} \right)^{-1} \right)$$

- Assim, um preditor para \mathbf{u}_j seria sua média condicional (à posteriori) ou seja, $\hat{\mathbf{u}}_j = \left(\mathbf{X}'_j \hat{\Sigma}_j^{-1} \mathbf{X}_j + \hat{\Psi}^{-1} \right)^{-1} \mathbf{X}'_j \hat{\Sigma}_j^{-1} (\mathbf{y} - \mathbf{Z}\hat{\gamma})$, em que $(\hat{\cdot})$ denota um dos estimadores vistos anteriormente (MVM ou MVR).

Cont.

- Medida de precisão de $\hat{\mathbf{u}}_j - \mathbf{u}_j$, $\text{Cov}(\hat{\mathbf{u}}_j - \mathbf{u}_j) = \Psi - \text{Cov}(\hat{\mathbf{u}}_j)$ (pois calcular somente $\text{Cov}(\hat{\mathbf{u}}_j)$ ignoraria a variabilidade contida em \mathbf{u}_j), em que (veja [Verbeke and Molenberghs \(2001\)](#))

$$\text{Cov}(\hat{\mathbf{u}}_j) = \Psi \mathbf{X}_j' \left(\mathbf{V}_j - \mathbf{V}_j \mathbf{Z}_j \left(\sum_{j=1}^J \mathbf{Z}_j' \mathbf{V}_j \mathbf{Z}_j \right)^{-1} \mathbf{Z}_j' \mathbf{V}_j \right) \mathbf{X}_j \Psi$$

Intervalos de Confiança

- Seja $\hat{\vartheta}$ o componente de interesse do vetor $\hat{\gamma}$ ou do vetor $\hat{\theta}$ e $\widehat{EP}(\hat{\vartheta})$ um estimador consistente (como aqueles apresentados) do respectivo erro-padrão.
- IC assintótico com coeficiente de confiança de τ

$$\hat{\vartheta} \pm z_{(1+\tau)/2} \widehat{EP}(\hat{\vartheta})$$

$$P(Z \leq z_{(1+\tau)/2}) = \frac{1+\tau}{2}$$

Testes de Hipótese

- Seja $\widehat{\Sigma}_\gamma$ um estimador consistente da matriz de covariâncias de $\widehat{\beta}$ (como aqueles apresentados).
- Desejamos testar $H_0 : \mathbf{C}\gamma = \mathbf{M}$ vs $H_1 : \mathbf{C}\gamma \neq \mathbf{M}$
- Podemos usar a seguintes estatística (do tipo Wald)

$$Q = (\mathbf{C}\widehat{\gamma} - \mathbf{M})' (\mathbf{C}\widehat{\Sigma}_\gamma\mathbf{C}')^{-1} (\mathbf{C}\widehat{\gamma} - \mathbf{M})$$

para $J, n_j, j = 1, \dots, J$ suficientemente grandes, temos que

$Q \sim \chi^2_{(r(\mathbf{C}), \delta)}$, em que

$$\delta = (\mathbf{C}\gamma - \mathbf{M})' (\mathbf{C}\Sigma_\gamma\mathbf{C}')^{-1} (\mathbf{C}\gamma - \mathbf{M})$$

Comentários

- Em relação aos testes de hipótese para θ , podemos proceder de modo análogo ao que fizemos para γ .
- Note, contudo, que podem existir três tipos de parâmetros em θ : parâmetros de variância, de correlação e de covariância. Para os parâmetros de variância, faz-se necessário testes mais específicos quando $M = 0$.
- Para outros detalhes, veja as referências.

Seleção de modelos: Teste da razão de verossimilhanças

- Seja $\hat{\vartheta}_i$ o estimador de máxima verossimilhança obtido sob o modelo i e $\tilde{\vartheta}_i$ sua respectiva estimativa.
- Denote por $L_i(\hat{\vartheta}_i)$ e $l_i(\hat{\vartheta}_i)$ o máximo da verossimilhança e da log-verossimilhança do modelo i , respectivamente, avaliados nos respectivos estimadores de MV (MVM/MVR), enquanto que $L_i(\tilde{\vartheta}_i)$ e $l_i(\tilde{\vartheta}_i)$ são os respectivos máximos avaliados nas estimativas de MV (MVM/MVR).
- No caso dos modelos mistos usa-se a log-verossimilhança marginal.
- Pode ocorrer problemas quando se usa a **verossimilhança completa** na seleção estrutural em MH (**Azevedo et al (2016)**).

Teste da razão de verossimilhanças (cont.)

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

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

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

γ = 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 algumas estatísticas de comparação de modelos, a seguir:

Estatísticas de comparação de modelos (cont.)

- Algumas estatísticas seguem: AIC (Akaike), BIC (Bayesiano), AICC(AIC corrigido), HQIC (Hannan e Quinn), CAIC(AIC consistente), SABIC (BIC ajustado pelo tamanho da amostra)

$$AIC_i = -2l_i(\tilde{\vartheta}_i) + 2k; BIC_i = -2l_i(\tilde{\vartheta}_i) + k \ln(n)$$

$$AICC_i = -2l_i(\tilde{\vartheta}_i) + \frac{2k(k+1)}{n-k-1}; HQIC_i = -2l_i(\tilde{\vartheta}_i) + 2k \ln(\ln(n))$$

$$CAIC_i = -2l_i(\tilde{\vartheta}_i) + k(\ln(n) + 1); SABIC = -2l_i(\tilde{\vartheta}_i) + k \ln\left(\frac{n+2}{24}\right)$$

Estatísticas de comparação de modelos (cont.)

- Em que $l_i(\tilde{\theta}_i)$ denota a log-verossimilhança do i -ésimo modelo avaliada em alguma estimativa (p.e. máxima verossimilhança), $k = q + r$ é o número de parâmetros e $n = \sum_{j=1}^J n_j$ é o número de observações.
- Portanto, o modelo que apresentar os menores valores, será o modelo “melhor ajustado” aos dados.