CE075 - Análise de Dados Longitudinais

Silva, J.L.P.

23 de setembro, 2019

Formulação em Dois Estágios

Formulação em dois Estágios do Modelo Linear Misto

No primeiro estágio as medidas longitudinais no i-ésimo indivíduo são modeladas como:

$$Y_i = Z_i \beta_i + \varepsilon_i,$$

em que Z_i covariáveis intra-indivíduo (ou tempo dependente) e $\varepsilon_i \sim N(0, \sigma^2 I_{n_i})$.

No segundo estágio, temos β_i aleatório (variando de indivíduo para indivíduo) tal que:

$$E(\beta_i) = A_i \beta,$$

Formulação em dois Estágios do Modelo Linear Misto

em que A_i $(q \times p)$ contém somente covariáveis que variam entre indivíduos (não dependente do tempo) e

$$Var(\beta_i) = D.$$

Desta forma,

$$\begin{aligned} Y_i &= Z_i(A_i\beta + b_i) + \varepsilon_i \\ &= X_i\beta + Z_i b_i + \varepsilon_i \end{aligned}$$

Ou seja, sob a restrição que

$$X_i = Z_i A_i$$

obtém-se o modelo de efeitos aleatórios.

Inferência Para o Modelo Misto

Inferência Para o Modelo Misto

Considere o modelo

$$Y_i = X_i\beta + Z_i b_i + \varepsilon_i,$$

em que, $b_i \sim N_q(0, D(\alpha))$ e $\varepsilon_{ij} \sim N(0, \sigma^2)$, b_i e ε_{ij} independentes.

Tem-se: p efeitos fixos e $\frac{q(q+1)}{2} + 1$ efeitos aleatórios.

Inferência Estatística para $\theta = (\beta, \alpha, \sigma^2)$:

1. Máxima Verossimilhança.
2. Máxima Verossimilhança Restrita.

Função de Verossimilhança

A função de verossimilhança é dada por:

$$
\begin{aligned}
L(\theta|y) &= \prod_{i=1}^{N} p(y_i|\theta) \\
&= \prod_{i=1}^{N} \int p(y_i, b_i|\theta) db_i \\
&= \prod_{i=1}^{N} \int p(y_i|b_i, \theta) p(b_i|\theta) db_i
\end{aligned}
$$

em que, $p(y_i|b_i, \theta) \sim N_{n_i}(X_i\beta + Z_i b_i, \sigma^2 I_{n_i})$ e $p(b_i|\theta) \sim N_q(0, D)$.

Observações

1. O EMV É obtido usando verossimilhança perfilada e iterações via algoritmo EM e/ou Newton-Raphson. Detalhes numéricos podem ser encontrados em Pinheiro e Bates (2000), Cap. 2.
2. O EMVR também pode ser obtido através de

$$l^*(\theta) = l(\theta) + termo.$$

3. A função `lme` do `R` fornece EMVR e EMV usando um enfoque híbrido (EM + Newton-Raphson). Esta função é de autoria de Pinheiro e Bates.
4. EMV e EMVR têm assintoticamente as propriedades usuais de um estimador de máxima verossimilhança (consistência e normalidade).

Avaliação dos Componentes de Variância

1. Número de componentes é igual a $\frac{q(q+1)}{2} + 1$ em que q é o número de efeitos aleatórios no modelo.
2. Muitas situações envolvem $q = 2$ (intercepto e inclinação aleatórios) e portanto:
$$\frac{2(2+1)}{2} + 1 = 4,$$
que permite termos heterogeneidade de variâncias e covariâncias pois ficam em função do tempo.
3. A escolha da "melhor" estrutura de variância-covariância pode ser realizada utilizando o teste da RMVR. Estes testes, usualmente, são na fronteira do espaço de parâmetros. Neste caso, a estatística da RMVR não tem, sob H_0, uma distribuição qui-quadrado.

Dist. da Estatística da RMVR sob H_0

1. A distribuição neste caso é uma mistura (50:50) de dist. qui-quadrado. Ou seja, por exemplo, para $H_0 : \sigma_{\beta_1} = 0$

$$RMVR \sim 0.5\chi_q + 0.5\chi_{q+1}$$

2. Exemplo
   Modelo completo: $q = 2$ (intercepto e inclinação aleatórios)
   Modelo restrito: $q = 1$ (somente intercepto aleatório)
   Teste usual (errado): nível de significância: 5,99
   Teste correto:

$$RMVR \sim 0,5\chi_1 + 0,5\chi_2$$

   nível é 5,14 (Tabela, Apend. C, Fitzmaurice et al, 2004).
3. Proposta ad hoc: para testar a $0,05$, use o nível de $0,10$.

Predição e Interpretação dos Efeitos Aleatórios

Predição dos Efeitos Aleatórios

Objetivo: predizer perfis individuais ou identificar indivíduos acima ou abaixo do perfil médio.

Obs.: não dizemos estimar os efeitos pois os mesmos são aleatórios. Dizemos predizer os efeitos aleatórios.

Deseja-se:

$$\widehat{Y}_i = \widehat{E}(Y_i|b_i) = X_i\widehat{\beta} + Z_i\widehat{b}_i$$

e para tal necessita-se de $\widehat{b}_i$, o chamado Estimador BLUP ("Best Linear Unbiased Predictor") de b_i.

Predição dos Efeitos Aleatórios

No modelo linear misto,

- Y_i e b_i tem uma distribuição conjunta normal multivariada.

- Usando conhecidas propriedades da normal multivariada, temos que

$$E(b_i|Y_i, \widehat{\beta}) = DZ_i' Var(Y_i)^{-1}(Y_i - X_i\widehat{\beta})$$

- Usando os estimadores MVR dos componentes de variância,

$$\widehat{b}_i = \widehat{D}Z_i' \widehat{Var}(Y_i)^{-1}(Y_i - X_i\widehat{\beta})$$

o BLUP de b_i.

Predição dos Efeitos Aleatórios

Desta forma obtemos:

$$
\begin{aligned}
\widehat{Y}_i &= X_i\widehat{\beta} + Z_i\widehat{b}_i \\
&= X_i\widehat{\beta} + Z_i\widehat{D}Z_i'\widehat{Var}(Y_i)^{-1}(Y_i - X_i\widehat{\beta}) \\
&= X_i\widehat{\beta} + (Z_i\widehat{D}Z_i' + \widehat{R}_i - \widehat{R}_i)\widehat{Var}(Y_i)^{-1}(Y_i - X_i\widehat{\beta}) \\
&= (\widehat{R}_i\widehat{Var}(Y_i)^{-1})X_i\widehat{\beta} + (I_{n_i} - \widehat{R}_i\widehat{Var}(Y_i)^{-1})Y_i
\end{aligned}
$$

em que $R_i = Var(Y_i|b_i) = Var(\varepsilon)$.

Interpretação: média ponderada entre a média populacional $X_i\widehat{\beta}$ e o i-ésimo perfil observado. Isto significa que o perfil predito é "encolhido" na direção da média populacional.

Predição dos Efeitos Aleatórios

A quantidade de "encolhimento" depende da magnitude de R_i e $Var(Y_i)$.

- R_i: variação intra-indivíduo:
- $Var(Y_i)$: variação total (entre e intra-indivíduo).

Quando R_i é relativamente grande, e a variabilidade intra indivíduo é maior que a variabilidade entre indivíduo, mais peso é atribuído a $X_i\hat{\beta}$, a média populacional estimada, do que à resposta individual observada.

Por outro lado, quando a variabilidade entre indivíduos é grande em relação à variabilidade intra indivíduos, mais peso é dado à resposta observada Y_i.

Predição dos Efeitos Aleatórios

Finalmente, o grau de "encolhimento" em direção à média populacional também depende de n_i.

Em geral, há maior encolhimento em direção à curva media populacional quando n_i é pequeno.

Intuitivamente, isso faz sentido já que menos peso deve ser dado à trajetória observada do indivíduo quando menos dados estão disponíveis.