Análise de Dados Longitudinais
Aula 05.09.2018

José Luiz Padilha da Silva - UFPR
www.docs.ufpr.br/∼jlpadilha

Sumário

Modelo Linear Misto

Ideia:

- Os parâmetros da regressão variam de indivíduo para indivíduo explicando as fontes de heterogeneidade da população.
- Cada indivíduo tem a sua própria trajetória média e um subconjunto dos parâmetros de regressão são tomados como aleatórios.
- Efeitos fixos são compartilhados por todos os indivíduos e os aleatórios são específicos de cada um.

Modelo Linear Misto

- **Características:**

  1. Características populacionais β (fixos);
  2. Características individuais β_i ou b_i (aleatórios).

- **Efeito:**

  1. Média: $E(Y_i) = X_i\beta$
  2. Estrutura de Covariância: Efeito aleatório induz $Var(Y_i)$.

     Separa a variação entre indivíduos daquela intra indivíduos.
  3. Permite obter estimativa de trajetórias individuais no tempo.

Modelo Linear Misto - Simetria Composta

Exemplo: $Y_{ij} = \beta_{0i} + \beta t_{ij} + \varepsilon_{ij}$ (Intercepto aleatório).

- $\beta_{0i} \sim N(\beta_0, \sigma^2_{\beta_0})$.
- $\varepsilon_{ij} \sim N(0, \sigma^2)$.
- β_{0i} e ε_{ij} são independentes.

1. $Var(Y_{ij}) = \sigma^2 + \sigma^2_{\beta_0}$.
2. $Cov(Y_{ij}, Y_{ij'}) = \sigma^2_{\beta_0}$

Modelo Linear Misto - Inclinação aleatória

Exemplo: $Y_{ij} = \beta_{0i} + \beta_{1i} t_{ij} + \varepsilon_{ij}$ (Intercepto e inclinação aleatórios).

- $\beta_{0i} \sim N(\beta_0, \sigma^2_{\beta_0}), \quad \beta_{1i} \sim N(\beta_1, \sigma^2_{\beta_1}), \quad Cov(\beta_{0i}, \beta_{1i}) = \sigma_{\beta_{01}}$.
- $\varepsilon_{ij} \sim N(0, \sigma^2)$.
- $\beta' = (\beta_{0i}, \beta_{1i})$ e ε_{ij} são independentes.

1. $Var(Y_{ij}) = \sigma^2_{\beta_0} + \sigma^2_{\beta_1} t^2_{ij} + 2 t_{ij} \sigma_{\beta_{01}} + \sigma^2$.
2. $Cov(Y_{ij}, Y_{ij'}) = \sigma^2_{\beta_0} + t_{ij} t_{ij'} \sigma^2_{\beta_1} + (t_{ij} + t_{ij'}) \sigma_{\beta_{01}}$.

Vantagens

1. Predizer trajetórias individuais (ex: intercepto aleatório)

$$Y_{ij} = X_{ij}\beta + b_i + \varepsilon_{ij}$$

   Resposta Média populacional:

$$E(Y_{ij}) = X_{ij}\beta$$

   Resposta média para o i-ésimo indivíduo (trajetória):

$$E(Y_{ij}|b_i) = X_{ij}\beta + b_i.$$

2. Flexibilidade em acomodar estruturas não balanceadas

Forma Geral do Modelo Misto

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

em que:

$(\beta)_{p\times 1}$: efeitos fixos;
$(b_i)_{q\times 1}$: efeitos aletaórios.

e,

$$b_i \sim N_q(0, \Sigma) \text{ e } \varepsilon_{ij} \sim N(0, \sigma^2)$$

Sendo b_i e ε_{ij} independentes.

$$q \leq p \Rightarrow Z_i \text{ é um subconjunto de } X_i$$

Incluímos efeitos aleatórios somente para as covariáveis que variam com o tempo.

Característica do Modelo

1. Média Populacional ou Marginal

$$E(Y_i) = X_i\beta.$$

2. Média condicional ou específica por indivíduo

$$E(Y_i|b_i) = X_i\beta + Z_i b_i.$$

3. Covariância Marginal

$$Var(Y_i) = Z_i Var(b_i) Z_i' + \sigma^2 I_{n_i}.$$

4. Podemos assumir que $\varepsilon_i \sim N(0, R_i)$ mas o usual é tomar $R_i = \sigma^2 I_{n_i}$ e interpretá-lo como covariância condicional. Ou seja,

$$Var(Y_i/b_i) = R_i = \sigma^2 I_{n_i}.$$

Inferência para o Modelo Misto

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

em que,

$$b_i \sim N_q(0, \Sigma(\alpha)) \text{ e } \varepsilon_{ij} \sim N(0, \sigma^2),$$

b_i e ε_{ij} independentes.

Desta forma 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

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

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

$$p(b_i|\theta) \sim N_q(0, \Sigma)$$

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 $0,10$.

Formulação em dois Estágios do Modelo Linear Misto

1. Estágio 1

   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

   $\varepsilon_i \sim N(0, \sigma^2 I_{n_i})$.

2. Estágio 2

   β_i: aleatório (variando de indivíduo para indivíduo) tal que:

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

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

Formulação em dois Estágios do Modelo Linear Misto

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

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.

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

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

- Usando os estimadores MVR dos componentes de variância,

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

o BLUP de b_i.

Predição dos Efeitos Aleatórios

Desta forma obtemos:

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

em que $Var(\varepsilon_i) = R_i$.

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.

Interpretação dos Efeitos Aleatórios Preditos

Ou seja,

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

- R_i grande, mais peso em $X_i\widehat{\beta}$;
- $Var(b_i)$ grande, mais peso em Y_i;
- n_i pequeno, mais peso em $X_i\widehat{\beta}$.