

# MAT02035 - Modelos para dados correlacionados

## Modelos lineares de efeitos mistos

Rodrigo Citton P. dos Reis  
citton.padilha@ufrgs.br

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA  
DEPARTAMENTO DE ESTATÍSTICA

Porto Alegre, 2019

# Formulação de efeitos aleatórios em dois estágios

# Formulação em dois estágios

O modelo linear de efeitos mistos

$$Y_i = X_i\beta + Z_ib_i + \epsilon_i,$$

pode ser motivado por uma formulação **de efeitos aleatórios em dois estágios** do modelo .

- ▶ Algumas das principais ideias do modelo de efeitos mistos são melhor compreendidas considerando que o modelo é resultante de uma especificação em dois estágios.
- ▶ No entanto, esta formulação introduz algumas restrições desnecessárias no modelo.

# Formulação em dois estágios: Estágio 1

- ▶ No primeiro estágio, assume-se que os indivíduos têm cada um a sua própria e única trajetória de resposta média.

$$Y_i = Z_i\beta_i + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma^2 I_{n_i}).$$

- ▶ A ideia essencial do modelo do primeiro estágio é ajustar modelos de regressão em separado para os dados de cada indivíduo, mas com a condição que estas regressões envolvem as mesmas covariáveis,  $Z_i$ .

# Formulação em dois estágios: Estágio 1

- ▶ A matriz  $Z_i$  especifica como a resposta média do indivíduo muda ao longo do tempo e/ou como a resposta média muda com outras covariáveis **tempo-dependente** (idade, altura).
  - ▶ Por exemplo, pode ser assumido que a trajetória da resposta média é **linear, quadrática**, ou uma **função spline do tempo**.
  - ▶ **Exemplo:** trajetórias indivíduo-específicas que são lineares no tempo.

$$\begin{pmatrix} Y_{i1} \\ Y_{i2} \\ \vdots \\ Y_{in_i} \end{pmatrix} = \begin{pmatrix} 1 & t_{i1} \\ 1 & t_{i2} \\ \vdots & \vdots \\ 1 & t_{in_i} \end{pmatrix} \begin{pmatrix} \beta_{1i} \\ \beta_{2i} \end{pmatrix} + \begin{pmatrix} \epsilon_{i1} \\ \epsilon_{i2} \\ \vdots \\ \epsilon_{in_i} \end{pmatrix}.$$

- ▶ **Exercício:** repita o exemplo para o caso de trajetórias quadráticas no tempo.

## Formulação em dois estágios: Estágio 2

- ▶ Assumimos que os  $\beta_i$  **são aleatórios**:  $\beta_i \sim D(E[\beta_i], \text{Cov}(\beta_i))$ .
- ▶ A média e a covariância de  $\beta_i$  são os parâmetros populacionais que são modelados no segundo estágio.
  - ▶ A variação de  $\beta_i$  de um indivíduo para outro é modelada como uma função de um conjunto de covariáveis **tempo-invariante** (grupo de tratamento).
- ▶ Em particular,

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

em que  $A_i$  é uma matriz  $q \times p$  de um conjunto de covariáveis tempo-invariante, e

$$\text{Cov}(\beta_i) = G.$$

- ▶ A especificação de um modelo para a média e a covariância de  $\beta_i$  completa o segundo estágio da formulação do modelo.

## Formulação em dois estágios: Estágio 2

- Considere o exemplo hipotético de um estudo que compara dois grupos: **tratamento** ( $\text{Grupo}_i = 1$ ) e **controle** ( $\text{Grupo}_i = 0$ ). Assumimos que as mudanças indivíduo-específicas na resposta média são lineares, ou seja, o primeiro estágio do modelo é dado por

$$\begin{pmatrix} Y_{i1} \\ Y_{i2} \\ \vdots \\ Y_{in_i} \end{pmatrix} = \begin{pmatrix} 1 & t_{i1} \\ 1 & t_{i2} \\ \vdots & \vdots \\ 1 & t_{in_i} \end{pmatrix} \begin{pmatrix} \beta_{1i} \\ \beta_{2i} \end{pmatrix} + \begin{pmatrix} \epsilon_{i1} \\ \epsilon_{i2} \\ \vdots \\ \epsilon_{in_i} \end{pmatrix}.$$

## Formulação em dois estágios: Estágio 2

- ▶ No segundo estágio podemos permitir que  $\beta_i$  dependa do grupo de tratamento

$$E(\beta_{1i}) = \beta_1 + \beta_3 \text{Grupo}_i,$$

$$E(\beta_{2i}) = \beta_2 + \beta_4 \text{Grupo}_i.$$

- ▶ Neste modelo,  $\beta_1$  é o intercepto médio no grupo controle, enquanto que  $\beta_1 + \beta_3$  é o intercepto médio do grupo tratamento.
  - ▶  $\beta_3$  representa a **diferença** do grupo tratamento **no intercepto médio**.
- ▶ Similarmente,  $\beta_2$  é a inclinação média, ou taxa de mudança na resposta média ao longo do tempo, no grupo controle, enquanto que  $\beta_2 + \beta_4$  é a inclinação média no grupo tratamento.
  - ▶  $\beta_4$  tem a interpretação em termos de uma diferença do grupo tratamento na inclinação média.



## Formulação em dois estágios: Estágio 2

- ▶ Neste modelo a matriz de delineamento  $A_i$  de covariáveis invariantes no tempo tem a seguinte forma:

$$A_i = \begin{pmatrix} 1 & 0 & \text{Grupo}_i & 0 \\ 0 & 1 & 0 & \text{Grupo}_i \end{pmatrix}.$$

## Formulação em dois estágios: Estágio 2

Assim, para o **grupo controle**, o modelo para média é

$$E \begin{pmatrix} \beta_{1i} \\ \beta_{2i} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{pmatrix} = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix};$$

similarmente, para o **grupo tratamento**, o modelo para média é

$$E \begin{pmatrix} \beta_{1i} \\ \beta_{2i} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{pmatrix} = \begin{pmatrix} \beta_1 + \beta_3 \\ \beta_2 + \beta_4 \end{pmatrix}.$$

## Formulação em dois estágios: Estágio 2

- ▶ Também é assumido que a variação residual em  $\beta_i$ , que não pode ser explicada pelo efeito de grupo, é

$$G = \begin{pmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{pmatrix},$$

em que  $g_{11} = \text{Var}(\beta_{1i})$ ,  $g_{22} = \text{Var}(\beta_{2i})$  e  $g_{12} = g_{21} = \text{Cov}(\beta_{1i}, \beta_{2i})$ .

- ▶  $g_{11}$  é a variância de  $\beta_{1i}$ , depois de ajustar pelo efeito de grupo de tratamento.

# Formulação em dois estágios

► Reescrevendo

$$\beta_i = A_i\beta + b_i,$$

em que  $b_i$  tem uma distribuição multivariada com média zero e matriz de covariância  $G$ , podemos combinar os dois componentes do modelo em dois estágios, obtemos

$$\begin{aligned} Y_i &= Z_i\beta_i + \epsilon_i \\ &= Z_i(A_i\beta + b_i) + \epsilon_i \\ &= (Z_iA_i)\beta + Z_ib_i + \epsilon_i \\ &= X_i\beta + Z_ib_i + \epsilon_i, \end{aligned}$$

em que  $X_i = Z_iA_i$ .

# Formulação em dois estágios: comentários

- ▶ Embora este modelo seja bastante similar àquele apresentado anteriormente, há uma importante diferença.
- ▶ O modelo de dois estágios impõe uma restrição na escolha da matriz de delineamento dos efeitos fixos, que requer a estrutura  $X_i = Z_i A_i$ , em que  $A_i$  contém apenas covariáveis invariantes no tempo e  $Z_i$  contém apenas covariáveis variantes no tempo.
- ▶ Isso implica que **qualquer covariável tempo-dependente** deve ser especificada como efeito aleatório, o que é uma restrição desnecessária e, em algumas situações, inconveniente!

# Formulação em dois estágios: comentários

- ▶ Por outro lado, uma estrutura simples para a covariância impõe uma estrutura bastante simples para a média!
- ▶ Vimos que uma covariância simetria composta é obtida do modelo com interceptos aleatórios,

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

sendo  $Z_i$  um vetor  $n_i \times 1$  de “uns”.

- ▶ Tal modelo impede qualquer dependência da resposta no tempo, isto é,

$$E(Y_i) = (Z_i A_i)\beta$$

não pode depender do tempo, pois o tempo, como variável intra-indivíduo, não foi incluído em  $Z_i$  no primeiro estágio.

# Inferência para o modelo linear de efeitos mistos

# Inferência para o modelo misto

Considere o modelo

$$Y_i = X_i\beta + Z_ib_i + \epsilon_i,$$

em que,  $b_i \sim N_q(0, G(\alpha))$  e  $\epsilon_{ij} \sim N(0, \sigma^2)$ ,  $b_i$  e  $\epsilon_{ij}$  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.



# Inferência para o modelo misto

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

- Note que  $p(y_i|\theta) \sim N_{n_i}(X_i\beta, Z_i G Z_i' + \sigma^2 I_{n_i})$ .

## **Escolha entre modelos de covariância de efeitos aleatórios**

## Escolha entre modelos de covariância

- ▶ Embora o modelo lineares de efeitos mistos assume que as respostas longitudinais dependem em uma combinação dos efeitos populacionais e indivíduo-específicos, quando tomamos a média com respeito a distribuição dos efeitos aleatórios

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

e a covariância entre as respostas tem a estrutura distinta de efeitos aleatórios

$$\text{Cov}(Y_i) = Z_i G Z_i' + \sigma^2 I_{n_i}.$$

## Escolha entre modelos de covariância

- ▶ Da perspectiva de modelar a covariância, a estrutura de efeitos aleatórios é atraente porque o **número de parâmetros de covariância**,  $q \times (q + 1)/2 + 1$ , é o mesmo, independentemente do número e do momento das ocasiões de medição.
- ▶ Em muitas aplicações, será suficiente incluir apenas interceptos aleatórios e inclinações para o tempo (um total de  $2 \times (2 + 1)/2 + 1 = 4$  parâmetros de covariância), **permitindo** assim a **heterogeneidade nas variâncias e correlações** que podem ser expressas **como funções do tempo**.
- ▶ Em outras aplicações, uma estrutura de efeitos aleatórios mais complexa pode ser necessária.

## Escolha entre modelos de covariância

- ▶ Na escolha de um modelo para a covariância, muitas vezes será interessante **comparar** dois modelos aninhados, um com  $q$  efeitos aleatórios correlacionados, outro com  $q + 1$  efeitos aleatórios correlacionados.
- ▶ A diferença no número de parâmetros de covariância entre esses dois modelos é  $q + 1$ , pois há uma variância adicional e  $q$  covariâncias adicionais no modelo “completo”.
- ▶ Conforme mencionado em aulas anteriores, o **teste da razão de verossimilhança** fornece um método válido para **comparar modelos aninhados** para a covariância.
- ▶ No entanto, em certos casos, a distribuição nula usual para o teste da razão de verossimilhança não é mais válida.

## Escolha entre modelos de covariância

- ▶ Estes testes, usualmente, são na **fronteira do espaço de parâmetros**.
  - ▶ Neste caso, a estatística da RV não tem, sob  $H_0$ , uma distribuição qui-quadrado.
- ▶ A distribuição neste caso é uma mistura de distribuições qui-quadrado.
  - ▶ Ou seja, por exemplo, para  $H_0 : \sigma_{b_2} = 0$

$$RV \sim 0.5\chi_q + 0.5\chi_{q+1}.$$

### 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:  $RV \sim 0.5\chi_1 + 0.5\chi_2$  nível é 5,14 (**Tabela, Apend. C, Fitzmaurice et al.**).

# Predição de efeitos aleatórios

# Predição de efeitos aleatórios

- ▶ **Objetivo:** prever perfis individuais ou identificar indivíduos acima ou abaixo do perfil médio.
- ▶ Deseja-se:

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

e para tal é necessário  $\hat{b}_i$ , o chamado Estimador BLUP (“Best Linear Unbiased Predictor”) de  $b_i$ .



# Predição de 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

$$E(b_i | Y_i, \hat{\beta}) = GZ_i' \Sigma_i^{-1} (Y_i - X_i \hat{\beta})$$

- ▶ Usando as estimativas de máxima verossimilhança dos componentes de variância,

$$\hat{b}_i = \hat{G}Z_i' \hat{\Sigma}_i^{-1} (Y_i - X_i \hat{\beta}),$$

o BLUP de  $b_i$ .

- ▶ (Abordagem **empirical Bayes**)

# Predição de efeitos aleatórios

$$\hat{Y}_i = X_i \hat{\beta} + Z_i \hat{b}_i = (\hat{R}_i \hat{\Sigma}_i^{-1}) X_i \hat{\beta} + (I_{n_i} - \hat{R}_i \hat{\Sigma}_i^{-1}) Y_i,$$

em que  $\text{Var}(\epsilon_i) = R_i$ , e

$$\hat{\Sigma}_i \hat{\Sigma}_i^{-1} = I_{n_i} = (Z_i \hat{G} Z_i' + \hat{R}_i) \hat{\Sigma}_i^{-1} = Z_i \hat{G} Z_i' \hat{\Sigma}_i^{-1} + \hat{R}_i \hat{\Sigma}_i^{-1}.$$

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

# Predição de efeitos aleatórios

- ▶ A quantidade de “encolhimento” (**shrinkage**) depende da magnitude de  $R_i$  e  $\Sigma_i$ .
  - ▶  $R_i$ : variância intra-indivíduo;
  - ▶  $\Sigma_i$ : variância total (entre e intra-indivíduo).
- ▶ Quando  $R_i$  é relativamente grande, e a variabilidade intra indivíduo é maior que a variabilidade entre indivíduos, 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 de 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 média 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.

# Avisos

# Avisos

- ▶ **Próxima aula (14/11):** Modelos lineares de efeitos mistos - exemplos e implementação computacional.
- ▶ **Para casa:** ler o Capítulo 8 do livro “**Applied Longitudinal Analysis**”.
  - ▶ Caso ainda não tenha lido, leia também os Caps. 1, 2, 3, 4, 5, 6 e 7.

Bons estudos!



**Instituto de**  
MATEMÁTICA  
E ESTATÍSTICA

---

60 anos

**UFRRGS**