

Análise de Dados Longitudinais

Aula 15.08.2018

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

- 1 Estimador de mínimos quadrados generalizados
- 2 Exemplo

Estimador de Mínimos Quadrados Generalizados

Se V_0 é conhecido, pode-se encontrar o Estimador de Mínimos Quadrados (generalizados).

$$\hat{\beta}_{MQG} = (X' V^{-1} X)^{-1} X' V^{-1} Y$$

Ideia:

Toda matriz positiva definida pode ser escrita como

$$V = K K' \quad K \text{ é não singular (existe inversa de } K)$$

Redefina o modelo como $Z = B\beta + \eta$, em que:

$$Z = K^{-1} Y$$

$$B = K^{-1} X$$

$$\eta = K^{-1} \varepsilon$$

Estimador de Mínimos Quadrados Generalizados

Assim,

$$\begin{aligned} \text{Var}(Z) &= \text{Var}(K^{-1}\varepsilon) \\ &= K^{-1} \text{Var}(\varepsilon) K^{-1'} \\ &= \sigma^2 K^{-1} K K' (K')^{-1} \\ &= \sigma^2 I_{Nn} \end{aligned}$$

Desta forma, retornamos à condição de Mínimos Quadrados Ordinários.

Equações Normais

$$\begin{aligned}
 \varepsilon' \varepsilon &= (Z - B\beta)'(Z - B\beta) \\
 &= (Y - X\beta)' V^{-1} (Y - X\beta) \\
 &= \sum_{i=1}^N (Y_i - X_i\beta)' V_0^{-1} (Y_i - X_i\beta)
 \end{aligned}$$

Resolver o sistema de equações:

$$\begin{aligned}
 \partial \varepsilon' \varepsilon / \partial \beta &= 2X' V^{-1} (Y - X\beta) \\
 &= \sum_{i=1}^N X_i' V_0^{-1} (Y_i - X_i \hat{\beta}) = 0
 \end{aligned}$$

Então

$$\begin{aligned}
 \hat{\beta} &= (B'B)^{-1} B'Z \\
 &= (X'K^{-1'}K^{-1}X)^{-1} X'K^{-1'} \cdot K^{-1}Y \\
 &= (X'K^{-1'}K^{-1}X)^{-1} X'(KK')^{-1}Y \\
 &= (X'V^{-1}X)^{-1} X'V^{-1}Y
 \end{aligned}$$

e

$$Var(\hat{\beta}) = \sigma^2(X'V^{-1}X)^{-1}$$

Resumo: EMQG

Modelo Linear:

$$Y_i = X_i' \beta + \varepsilon_i$$

tal que $E(\varepsilon_i) = 0$, $Var(\varepsilon_i) = \sigma^2 V_0$ e V_0 matriz de correlação.

Restrição: homocedasticidade (não é necessário mas conveniente).

$$\hat{\beta}_{MQG} = (X' V^{-1} X)^{-1} X' V^{-1} Y, \quad \text{não depende de } \sigma^2,$$

V : verdadeira estrutura de covariância para Y_i

$$Var(\hat{\beta}_{MQG}) = \sigma^2 (B' B)^{-1} = \sigma^2 (X' V^{-1} X)^{-1}$$

O EMQG somente é válido se V for conhecida.

Modelo Marginal

Pergunta: Em situações reais V não é conhecido. O que devemos fazer?

- 1 Resposta Normal: Utilizar o Método de Máxima Verossimilhança (usual ou restrito) para estimar β e também os componentes de variância. Ou seja, os parâmetros da média e também da estrutura escolhida de covariância.
- 2 Sem especificar distribuição para a resposta: Investigar qual é o impacto ao utilizarmos W ao invés de V . Ideia de GEE. (W a princípio pode ser aquela mais adequada para modelar a estrutura de covariância dos dados.)

Estimador de Máxima Verossimilhança

Encontrar simultaneamente o estimador da média (β) e o estimador para os componentes de variância (σ^2, α).

Seja

$$Y_i \sim N_n(X_i\beta, \sigma^2 V_0(\alpha))$$

$$f(y_i|\beta, \sigma^2, \alpha, X_i) = \frac{1}{(2\pi)^{n/2} |V_0|^{1/2} (\sigma^2)^{n/2}} \exp \left\{ -\frac{1}{2\sigma^2} Q_i \right\}$$

em que

$$Q_i = (y_i - X_i\beta)' V_0^{-1} (y_i - X_i\beta)$$

Exemplo: Chumbo em Crianças

- Modelo Não-Estruturado para a média (intercepto comum):
(R: $y \sim \text{factor}(\text{tempo}) * \text{factor}(\text{grupo})$).
- Comparando estruturas para $\text{Var}(Y_i)$.
- Estimativas para média e erro-padrão para os coeficientes dos termos da interação.

Coeficiente	Independente		Simetria Composta		AR1		Não Estruturada	
	Est.	EP	Est.	EP	Est.	EP	Est.	EP
Linha base	-0,268	1,325	-0,268	1,325	-0,268	1,318	-0,268	1,326
1a semana	11,406	1,874	11,406	1,192	11,406	1,132	11,406	1,192
4a semana	8,824	1,874	8,824	1,192	8,824	1,446	8,824	1,121
6a semana	3,152	1,874	3,152	1,192	3,152	1,612	3,152	1,278

Exemplo: Chumbo em Crianças

- 1 Modelo Não-Estruturado para a média.
- 2 Algumas estruturas para $Var(Y_i)$.
- 3 Estimativas dos parâmetros da média (β) não mudam ao estruturarmos $Var(Y_i)$.
- 4 O mesmo não acontece com as estimativas dos erros-padrão. Observe a diferença, principalmente entre a independente e as demais.
- 5 Pela análise exploratória as formas independente e AR1 aparentemente não são adequadas para modelar estes dados.