

Análise de Dados Longitudinais

Aula 22.08.2018

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

Sumário

- 1 O estimador GEE
- 2 Método de estimação
- 3 Exemplo

O Estimador GEE

O estimador GEE para β é dado por:

$$\sum_{i=1}^N X_i' V_i^{-1}(\alpha)(y_i - X_i\beta) = 0,$$

em que α são os componentes de variância.

Usualmente tomamos:

$$V_i(\alpha) = A_i^{1/2} R_i A_i^{1/2}$$

em que $A_i(\alpha)$ é uma matriz diagonal com elementos $Var(Y_{ij})$ e $R_i(\alpha) = Corr(Y_i)$ (matriz de trabalho) é matriz de correlação.

Formas de Correlação de Trabalho

- *independência*,
⇒ dados longitudinais não correlacionados.
- *simetria composta*,
⇒ equivalente a um modelo linear misto com apenas o intercepto aleatório.
- *AR1*,
⇒ válida para medidas igualmente espaçadas no tempo;
- *não estruturada* estima todas as $n(n - 1)/2$ correlações de ***R***.
- Outras: banded, toeplitz, etc.

Variância do Estimador

- 1 *Naive* ou “baseada no modelo” - Viciada

$$\widehat{Var}(\hat{\beta}) = \left(\sum_{i=1}^N X_i' V_i(\hat{\alpha})^{-1} X_i \right)^{-1}.$$

- 2 *Robusta* ou “empírica” ou Sanduíche

$$\widehat{Var}(\hat{\beta}) = M_0^{-1} M_1 M_0^{-1},$$

em que

$$M_0 = \sum_{i=1}^N X_i' V_i(\hat{\alpha})^{-1} X_i,$$

$$M_1 = \sum_{i=1}^N X_i' V_i(\hat{\alpha})^{-1} (y_i - \hat{\mu}_i)(y_i - \hat{\mu}_i)' V_i(\hat{\alpha})^{-1} X_i.$$

Método de Estimação: GEE - Passos

- 1 Especificar a forma de $R(\alpha)$ e tomar

$$\hat{\beta} = (XX)^{-1} X' Y$$

- 2 Encontrar os resíduos $e_{ij} = Y_{ij} - X_{ij}\hat{\beta}$. Através dos resíduos é possível estimar α e σ^2 , componentes de V_i . Por exemplo, estimamos

$$\hat{\sigma}^2 = \frac{\sum_i \sum_j e_{ij}^2}{nN}.$$

- 3 Dado $V(\hat{\alpha})$, obtemos

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

Retornar ao passo 2 até a convergência.

Método de Estimação: GEE - Passos

Os parâmetros de associação α são obtidos dos resíduos, dependendo do modelo assumido para a associação.

Por exemplo, em um desenho balanceado, quando a associação é expressa em termos de correlações não estruturadas, $\alpha_{jk} = \text{Corr}(Y_{ij}, Y_{ik})$ pode ser estimado por:

$$\hat{\alpha}_{jk} = \left(\frac{1}{\hat{\sigma}^2 N} \right) \sum_i \sum_j e_{ij} e_{ik}.$$

Método de Estimação: GEE - Passos

Após a convergência estimar $Var(Y_i)$ e obter $\widehat{Var}(\hat{\beta})$:

$$\begin{aligned}\widehat{Var}(\hat{\beta}) &= (X' \hat{V}^{-1} X)^{-1} X' \hat{V}^{-1} \widehat{Var}(Y_i) \hat{V}^{-1} X (X' \hat{V}^{-1} X)^{-1} \\ &= \hat{M}_0^{-1} \hat{M}_1 \hat{M}_0^{-1}\end{aligned}$$

GEE - Observações

- 1 Este estimador de $\text{Var}(\hat{\beta})$ é chamado de estimador sanduíche (M_0^{-1} é o pão e M_1 é a carne)
- 2 Se tomarmos $V = I$, temos

$$\widehat{\text{Var}}(\hat{\beta}_I) = (X'X)^{-1} X' \widehat{\text{Var}}(Y_i) X (X'X)^{-1}$$

- 3 Se tomarmos $V = \text{Var}(Y_i)$,

$$\widehat{\text{Var}}(\hat{\beta}_V) = \hat{M}_0^{-1}$$

GEE - Características e Limitações

1 Vantagens/Características

- $\hat{\beta}$ é consistente mesmo que $Var(Y_i)$ for incorretamente especificada.
- Não é necessário especificar uma distribuição para Y_i .
- $Var(\hat{\beta})$ é adequadamente estimada pelo estimador sanduíche.

2 Limitações

- A robustez do estimador sanduíche é uma propriedade assintótica.
- A matriz de trabalho V_i deve ser especificada o mais próximo possível de $Var(Y_i)$ para obter eficiência/precisão para a estimação de β .

GEE - Características e Limitações

2 Continuação: Limitações

- O estimador GEE, $\hat{\beta}$ fica viciado na presença de dados perdidos se a matriz de trabalho não for corretamente especificada e o mecanismo de perda não for MCAR.
- Na maioria dos softwares σ^2 é tomado como sendo invariante no tempo. Ou seja, $\sigma_1^2 = \sigma_2^2 = \dots = \sigma_n^2 = \sigma^2$. Este fato é restritivo para analisar respostas contínuas.

Exemplo: Chumbo em Crianças - GEE

- Modelo Não-Estruturado para a média (intercepto comum):
(R: `y ~ factor(tempo)*factor(grupo)`).
- Comparando estruturas para $Var(Y_i)$.
- Estimativas para média e erro-padrão para os coeficientes que comparam os grupos nos quatro tempos.

Coeficiente	Independente		Simetria Composta		AR1		Não Estruturada	
	Est.	EP	Est.	EP	Est.	EP	Est.	EP
Linha base	-0,268	0,994	-0,268	0,994	-0,268	0,994	-0,268	0,994
1a semana	11,406	1,109	11,406	1,109	11,406	1,109	11,406	1,109
4a semana	8,824	1,141	8,824	1,141	8,824	1,141	8,824	1,141
6a semana	3,152	1,244	3,152	1,244	3,152	1,244	3,152	1,244