Análise de Dados Longitudinais
Aula 20.08.2018

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

Sumário

Flexibilizar Suposições - GEE

1. Investigar qual é o impacto ao utilizarmos erradamente W ao invés da estrutura correta V.
2. Investigar a possibilidade de não especificar distribuição para Y.
3. Princípio das Equações de Estimação Generalizadas (GEE).

EMQG

Supor que ao invés de V foi utilizada erradamente W. Ou seja,

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

Pergunta: Qual é o impacto na estimação de β se utilizarmos W ao invés de V?

Isto é,

- Qual é o vício de $\widehat{\beta}_W$?
- Qual é a $Var(\widehat{\beta}_W)$?

Vício e Variância

1. **Vício:**

$$\begin{aligned} E(\hat{\beta}_W) &= (X'W^{-1}X)^{-1}X'W^{-1}E(Y) \\ &= (X'W^{-1}X)^{-1}X'W^{-1}X\beta \\ &= \beta \end{aligned}$$

2. **Variância:**

$$\begin{aligned} Var(\hat{\beta}_W) &= Var\left[(X'W^{-1}X)^{-1}X'W^{-1}Y\right] \\ &= (X'W^{-1}X)^{-1}X'W^{-1}Var(Y)\left[W^{-1}X(X'W^{-1}X)^{-1}\right] \\ &= \sigma^2(X'W^{-1}X)^{-1}X'W^{-1}VW^{-1}X(X'W^{-1}X)^{-1} \end{aligned}$$

Especificação Incorreta de W

Pergunta: O que acontece ao especificarmos um W errado? (Ou seja, longe de V.)

- $\hat{\beta}_W$ é não-viciado para qualquer especificação de W;
- Por exemplo, se $W = I_{Nn}$

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

Observações:

- $\hat{\beta}_I = (X'X)^{-1}X'Y$ (EMQ) é não viciado.
- $Var(\widehat{\beta}_I) = \sigma^2(X'X)^{-1}$, é viciada.

Especificação Incorreta de W

Pergunta Quanto $Var(\hat{\beta}_I)$ é diferente de $Var(\hat{\beta}_{MQG})$?

Ou seja, quanto

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

é diferente de

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

Resposta Na maioria da vezes estes estimadores são bem próximos.

Exemplo (Diggle et al., p. 59):

$$
\begin{aligned}
N &= 10 \\
n &= 5 \qquad (t = -2, -1, 0, 1, 2)
\end{aligned}
$$

$W = I_{50}$

$$V_0 = [(1-\rho)I_5 + \rho 1_5 1_5']$$

Modelo: $Y_{ij} = \beta_0 + \beta_1 t_{ij} + \varepsilon_{ij}$

$$
X_{50,2} = \begin{pmatrix} 1 & -2 \\ 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ \vdots & \vdots \\ 1 & 2 \end{pmatrix}
$$

Exemplo (Diggle et al., p. 59):

Fazendo as contas:

$$X'X = \begin{pmatrix} 50 & 0 \\ 0 & 100 \end{pmatrix}$$

e

$$X'VX = \begin{pmatrix} 50(1+4\rho) & 0 \\ 0 & 100(1-\rho) \end{pmatrix}.$$

Desta forma,

$$Var(\hat{\beta}_I) = \sigma^2 (X'X)^{-1} X'VX (X'X)^{-1} = \sigma^2 \begin{pmatrix} 0.02(1+4\rho) & 0 \\ 0 & 0.01(1-\rho) \end{pmatrix}.$$

Exemplo (Diggle et al., p. 59):

Continuando as contas:

$$V_0^{-1} = (1-\rho)^{-1}\rho((1-\rho)(1+4\rho))^{-1}1_5 1_5'$$

e

$$X'V^{-1}X = \begin{pmatrix} 50(1+4\rho)^{-1} & 0 \\ 0 & 100(1-\rho)^{-1} \end{pmatrix}.$$

Desta forma,

$$Var(\hat{\beta}_{MQG}) = \sigma^2(X'V^{-1}X)^{-1} = \sigma^2 \begin{pmatrix} 0.02(1+4\rho) & 0 \\ 0 & 0.01(1-\rho) \end{pmatrix}.$$

Ou seja, neste caso $Var(\hat{\beta}_I) = Var(\hat{\beta}_{MQG})$.

Observação: Em várias situações a $Var(\hat{\beta}_I)$ é um estimador razoável para $Var(\hat{\beta}_{MQG})$.

Resumo

Assumindo o estimador de Mínimos Quadrados Ordinário $W = I_{Nn}$:

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

$$E(\hat{\beta}_I) = \beta$$

e sua variância fica usualmente bem estimada por:

$$Var(\hat{\beta}_I) = (X'X)^{-1}X'VX(X'X)^{-1}.$$

Precisamos de um estimador consistente para V!!

Estimador Consistente de V

Um estimador consistente para V é dado por:

$$\hat{V}_{0i} = (Y_i - X_i'\hat{\beta})(Y_i - X_i'\hat{\beta})'$$

$$\hat{V} = \begin{bmatrix} \hat{V}_{01} & 0 & \cdots & 0 \\ 0 & \hat{V}_{02} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \hat{V}_{0N} \end{bmatrix}_{Nn \times Nn}$$

Obs. O parâmetro σ^2 foi absorvido em V.

Equações de Estimação Generalizadas (GEE)

Proposto por Liang e Zeger (1986) para dados correlacionados.

Requer apenas a especificação correta da estrutura de média das variáveis respostas, sem fazer qualquer suposição distribucional.

Especificamos:

1. $E(\mathbf{Y}_i) = \mathbf{X}_i\boldsymbol{\beta} = \boldsymbol{\mu}_i$, e
2. matriz de correlação "de trabalho" das medidas repetidas, $\boldsymbol{R_i(\alpha)}$.

GEE gera estimadores consistentes e assintoticamente normais para β, mesmo com má especificação $\boldsymbol{R_i(\alpha)}$.

O Estimador GEE - Motivação

Uma motivação para o enfoque GEE vem dos estimadores de MQG que minimiza a função objetivo:

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

O estimador de $\boldsymbol{\beta}$, específico para o modelo linear, é a solução de

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

que produz, resolvendo para β,

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

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*,
  $\Rightarrow$ dados longitudinais não correlacionados.
- *simetria composta*,
  $\Rightarrow$ equivalente a um modelo linear misto com apenas o intercepto aleatório.
- *AR1*,
  $\Rightarrow$ válida para medidas igualmente espaçadas no tempo;
- *não estruturada* estima todas as $n(n-1)/2$ correlações de $\boldsymbol{R}$.
- 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

$$\begin{aligned} M_0 &= \sum_{i=1}^{N} X_i' V_i(\hat{\alpha})^{-1} X_i, \\ M_1 &= \sum_{i=1}^{N} X_i' V_i(\hat{\alpha})^{-1} (y_i - \hat{\mu}_i)(y_i - \hat{\mu}_i)' V_i(\hat{\alpha})^{-1} X_i. \end{aligned}$$