Análise de Dados Longitudinais
Modelo Marginal

Enrico A. Colosimo/UFMG

Modelos Marginais para Dados Longitudinais

1. Modelar a resposta média $E(Y)$.

2. Modelar a Estrutura de Variância-Covariância $Var(Y_i), i = 1, \ldots N$.

3. Assumir uma distribuição (normal) para a resposta (dispensável).

Modelo Linear Geral p- covariáveis

$$Y_{ij} = \beta_1 X_{ij1} + \beta_2 X_{ij2} + \cdots + \beta_p X_{ijp} + \varepsilon_{ij}; \quad i = 1, \ldots, N; \ j = 1, \ldots, k,$$

em que $X_{ij1} = 1$. Escrevendo em forma matricial.

$$Y_i = \begin{pmatrix} Y_{i1} \\ Y_{i2} \\ \vdots \\ Y_{ik} \end{pmatrix} = X_i \beta + \varepsilon_i$$

em que X_i tem dimensão $k \times p$ e β é um vetor p-variado.

Estrutura de Variância-Covariância

$$Var(Y_i) = \sigma^2 V_0 \quad \text{(supondo homocedasticiadade)}$$

e desta forma V_0 é a matriz de correlação de Y_i.

Como as unidades formam uma amostra aleatória da população temos que:

$$V = \begin{pmatrix} V_0 & 0 & \cdots & 0 \\ 0 & V_0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & V_0 \end{pmatrix}$$

Modelo para a Média

$$E(Y_{ij}) = \beta_1 X_{ij1} + \beta_2 X_{ij2} + \cdots + \beta_p X_{ijp}$$

- A média é uma combinação linear dos parâmetros.
- A estruturação da média deve ser baseada na análise exploratória dos dados ou informação histórica.
- Usamos termos polinomiais e splines para o comportamento temporal.

Exemplos de Formas para V_0

1 Simetria Composta

$$V_0 = [(1-\rho)I_k + \rho 1_k 1_k']$$

em que

$$1_k 1_k' = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix}_{k\times k}$$

Ou seja, $1_k' = (1, 1, \cdots, 1)$ que é um vetor de k 1's.

Justificativa

$$Y_{ij} = X_{ij}'\beta + U_i + \varepsilon_{ij}$$

O intercepto é o único termo com variação aleatória.
A diferença entre os indivíduos está explicada pelo intercepto aleatório:

$$\begin{aligned} U_i &\sim N(0, \sigma_u^2) \\ \varepsilon_{ij} &\sim N(0, \sigma^2), \end{aligned}$$

em que U_i e ε_{ij} são independentes.

2 Correlação Exponencial (AR(1)):

$$V_0 = \begin{bmatrix} 1 & V_{12} & V_{13} & \cdots & V_{1k} \\ & 1 & V_{23} & \cdots & V_{2k} \\ & & \ddots & & \\ & & & & 1 \end{bmatrix}$$

$$V_{il} = \rho^{|t_j - t_l|}$$

ou

$$V_{il} = \rho^{|j-l|}, \qquad \text{se for balanceado e equidistante}$$

Justificativa

$$\varepsilon_{ij} = \rho\varepsilon_{ij-1} + Z_{ij}$$
$$Z_{ij} \sim N(0, \sigma^2(1-\rho^2)),$$

em que ε_{ij} e Z_{ij} são independentes.

Então

$$Var(\varepsilon_{ij}) = \rho^2\sigma^2 + \sigma^2(1-\rho^2) = \sigma^2$$
$$Cov(\varepsilon_{ij}, \varepsilon_{ij-1}) = Cov(\rho\varepsilon_{ij-1} + Z_{ij}, \varepsilon_{ij-1}) = \rho\sigma^2 \quad \text{e}$$

para lags maiores que 1,

$$Cov(\varepsilon_{ij}, \varepsilon_{ij-l}) = \rho^l\sigma^2$$

Observações com relação ao Exponencial AR(1)

1. Este modelo é conhecido por Auto-regressivo de primeira ordem: AR(1).
2. É possível usa-lo para modelar desenhos não balanceados. Segue a mesma ideia do variograma.

Exemplos

3 Toeplitz: Extensão do AR(1)

$$V_0 = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-2} \\ \rho_2 & \rho_1 & \ddots & & \rho_{k-3} \\ & & & & 1 \end{bmatrix}$$

De uma forma geral,

$$Cov(Y_{ij}, Y_{ij+l}) = \rho_l$$

Exemplos

4 Banded:

$$V_0 = \begin{bmatrix} 1 & \rho_1 & 0 & \cdots & 0 \\ \rho_1 & 1 & \rho_1 & \cdots & 0 \\ 0 & \rho_1 & \ddots & & 0 \\ & & & & 1 \end{bmatrix}$$

Caso particular do Toeplitz quando fazemos ,

$$\rho_2 = \rho_3 = \cdots \rho_{k-1} = 0$$

5 Modelos Híbridos:

$$Cov(Y_i) = \sigma_1^2 V_1 + \sigma_2^2 V_2.$$

Por exemplo: V_1 é de simetria composta e V_2 é AR(1).

Se nenhuma forma estruturada for adequada para um particular conjunto de dados, devemos utilizar a forma não-estruturada.

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

$V = KK'$ K é não singular(existe inversa de K)

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

$$\begin{aligned} Z &= K^{-1}Y \\ B &= K^{-1}X \\ \eta &= K^{-1}\varepsilon \end{aligned}$$

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

Desta forma, retornamos a 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:

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

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 distribuiçao 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

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

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

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: $y \sim$ factor(tempo)*factor(grupo)).
- Comparando estruturas para $Var(Y_i)$.
- 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.

Observação no banco: Chumbo em Crianças

- Ao usarmos uma matriz de desenho ”ortogonal”, o OLS coincide com o GLS.

- Modelo Não-Estruturado para a média (sem intercepto): (R: y   factor(tempo) + factor(tempo)*factor(grupo)-1).

- Neste caso (deve ser mostrado analiticamente)

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

- Observe que este modelo é mais intuitivo para a comparação dos grupos em cada tempo.

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	1,325	11,406	1,325	11,406	1,318	11,406	1,326
4a semana	8,824	1,325	8,824	1,325	8,824	1,318	8,824	1,326
6a semana	3,152	1,325	3,152	1,325	3,152	1,318	3,152	1,326

Crianças - Transmissão Vertical

- Modelo quadrático para a média com termos de interação.
- Algumas formas para a $Var(Y_i)$: exponencial, simetria composta.
- Modelo para média com 9 termos (interceptos diferentes)
- Resultados para os quatro termos da interação.

Coeficiente	Independente		Exponencial		Simetria Composta	
	Est.	EP	Est.	EP	Est.	EP
Idade:grupo	$-0{,}164$	0,041	$-0{,}160$	0.057	$-0{,}142$	0,027
Idade2:grupo	0,020	0,008	0,017	0,008	0,018	0,005
Idade:sexo	0,046	0,037	0,165	0,052	0,100	0,025
Idade2:sexo	$-0{,}014$	0,007	$-0{,}020$	0,008	$-0{,}015$	0,005

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 distribuiçao para Y.
3. Princípio das Equações de Estimação Generalizadas (GEE).

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 $Var(\widehat{\beta}_W)$?

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:**

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

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 $W = I_{Nk}$

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

Observações:

- $\hat{\beta}_I = (X'X)^{-1}X'Y$ (Estimador de Mínimos Quadrados Ordinários) é não viciado.
- No entanto,

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

é viciada.

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):

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

$W = I_{50}$

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

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

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

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,

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

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 $W = I_{Nk}$:

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

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

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

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

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

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. $E(\mathbf{Y}_i) = \mathbf{X}_i\beta = \boldsymbol{\mu}_i$, 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 $R_i(\phi) = Corr(Y_i)$ (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 $k(k-1)/2$ 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

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

Método de Estimação: GEE - Passos

1. Escolher $R(\phi)$: matriz de trabalho e usualmente assumimos $A = \sigma^2 I_k$ (homocedasticidade)) e $\alpha = (\sigma, \phi)$.

2. Dado estimativas para ϕ e σ, obtemos $V(\hat{\alpha})$ e obtemos:

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

   Obs. Inicializar o processo iterativo com $R(\phi) = I_k$.

3. Encontrar os resíduos: $e_{ij} = Y_{ij} - X_{ij}\widehat{\beta}$. A partir dos resíduos é possível estimar

$$\widehat{\sigma}^2 = \frac{\sum_i \sum_j e_{ij}^2}{kN}$$

   e também os outros componentes de variância ϕ. Retornar ao passo 2 até a convergência.

Método de Estimação: GEE - Passos

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

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

Obs. A estimativa de ϕ é baseada nos resíduos. Por exemplo, em um desenho balanceado a desenho não estruturado é estimado por:

$$\hat{\phi}_{jk} = \frac{1}{\hat{\sigma}N}\sum_{i=1}^{N} e_{ij}e_{ik}$$

GEE - Observações

1. Este estimador de $Var(\widehat{\beta})$ é chamado de estimador sanduíche (M_0^{-1} é o pão e M_1 é a carne)

2. Se tomarmos $V = I$, temos

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

3. Se tomarmos $V = Var(Y_i)$,

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

GEE - Características e Limitações

1. Vantagens/Características
   - $\widehat{\beta}$ é consistente mesmo que $Var(Y_i)$ for incorretamente especificada.
   - Não é necessário especificar uma distribuição para Y_i.
   - $Var(\widehat{\beta})$ é adequadamente estimada pelo estimador sanduíche.
2. Limitações
   - Desenho desbalanceado é uma restrição para a estimação usando GEE, especialmente para o estimador sanduíche.
   - 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.

2 Continuação: Limitações

- O estimador GEE, $\widehat{\beta}$ 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, $\sigma_1^2 = \sigma_2^2 = \cdots, \sigma_k^2 = \sigma^2$. 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	$-0{,}268$	1,318	$-0{,}268$	1,326
1a semana	11,406	1,109	11,406	1,109	11,406	1,132	11,406	1,192
4a semana	8,824	1,141	8,824	1,141	8,824	1,446	8,824	1,121
6a semana	3,152	1,244	3,152	1,244	3,152	1,612	3,152	1,278

Crianças - Transmissão Vertical - GEE

- Modelo quadrático para a média com termos de interação.
- Algumas formas para a $Var(Y_i)$: exponencial, simetria composta.
- Modelo para média com 9 termos (interceptos diferentes)
- Resultados para os quatro termos da interação.

Coeficiente	Independente		Simetria Composta	
	Est.	EP	Est.	EP
Idade:grupo	$-0{,}164$	0,059	$-0{,}142$	0,057
Idade2:grupo	0,020	0,011	0,018	0,008
Idade:sexo	0,046	0,050	0,100	0,047
Idade2:sexo	$-0{,}014$	0,009	$-0{,}015$	0,007

Obs. As estimativas são as mesmas do gls-normal e o erro-padrão fica inflacionado.

Revisão - Teoria de Verossimilhança

Considere $Y_1, Y_2, \ldots, Y_n$ respostas iid de uma população $f(y;\theta)$. Então a função de verossimilhança para θ é dada por

$$L(\theta/y) = \prod_{i=1}^{n} f(y_i/\theta),$$

em que θ é um vetor de p-parâmetros.

O EMV (Estimador de Máxima Verossimilhança) é aquele $\widehat{\theta}$ que maximiza $L(\theta/y)$ ou, de forma equivalente, $l(\theta/y) = log(L(\theta/y))$ no espaço de parâmetros de θ.

Revisão - Teoria de Verossimilhança: Função Escore

$$S(\theta) = \frac{\partial l(\theta/y)}{\partial\theta},$$

que é p-dimensional.

O EMV é a solução do sistema de equações determinado pela função escore:

$$S(\widehat{\theta}) = 0.$$

Propriedade importante: $E(S(\theta)) = 0$.

Revisão - Teoria de Verossimilhança: Medida de Incerteza

$$\begin{aligned}\Im(\theta) &= Var(S(\theta))\\ &= \mathrm{E}(S(\theta)^2)\\ &= -\mathrm{E}\left(\frac{\partial^2 l(\theta)}{\partial\theta\partial\theta'}\right),\end{aligned}$$

que é uma matriz $p \times p$ chamada de Informação de Fisher.

A variância assintótica de $\widehat{\theta}$ é

$$Var(\hat{\theta}) = \Im(\theta)^{-1}$$

que é estimada avaliando θ em $\widehat{\theta}$.

Revisão - Teoria de Verossimilhança: Medida de Incerteza

Usualmente é díficil encontrar o valor esperado na distribuição de Y. No entanto, podemos utilizar qualquer estimador consistente de $\Im$. Usamos a matriz de informação observada

$$I(\theta) = -\left(\frac{\partial^2 l(\theta)}{\partial\theta\partial\theta'}\right),$$

que é consistente para $\Im(\theta)$. Ou seja,

$$Var(\widehat{\theta}) \approx I(\theta)^{-1}.$$

Obs. O resultado é verdadeiro para qualquer estimador consistente de $\Im$.

Revisão - Teoria de Verossimilhança: Estatísticas

1. Wald
2. Razão de Verossimilhança
3. Escore

Estimador de Máxima Verossimilhança

A função de Verossimilhança:

$$L(\beta, \sigma^2, \alpha) = \prod_{i=1}^{N} f(y_i|\beta, \sigma^2, \alpha, X_i)$$

e a função de log-verossimilhança:

$$\begin{aligned} l(\beta, \sigma^2, \alpha) = & - \frac{kN}{2}\left[log(2\pi) + log(\sigma^2)\right] - \frac{N}{2} log(|V_0|) - \\ & - \frac{1}{2\sigma^2}\sum_{i=1}^{N}\left\{(y_i - X_i\beta)' V_0^{-1}(y_i - X_i\beta)\right\} \end{aligned}$$

Observações:

1. O vetor de parâmetros β somente aparece no último termo;
2. Se V_0 e σ^2 forem fixos , o estimador de β consiste em minimizar :

$$\sum_{i=1}^{N}(y_i - X_i\beta)' V_0^{-1}(y_i - X_i\beta)$$

   cuja solução é:

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

3. Sob heterocedasticidade, as variâncias são absorvidas em V_0. Ou seja, $Var(Y_i) = V_0$ e V_0 não é mais a matriz de correlação.

Usando verossimilhança perfilada para obter o EMV de β, σ^2 e α:

1. fixamos inicialmente V_0 e σ^2 e obtemos:

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

2. Em seguida substuimos β por $\hat{\beta}(\alpha, \sigma^2)$ em

$$\begin{aligned} l(\hat{\beta}, \alpha, \sigma^2) & \propto & \frac{-N}{2}\left(k\log(\sigma^2) + \log(|V_0|)\right) - \\ & - & \frac{1}{2\sigma^2}\sum_{i=1}^{N}\left[(y_i - X_i'\hat{\beta})V_0^{-1}(y_i - X_i'\hat{\beta})\right] \end{aligned}$$

Verossimilhança Perfilada

Tomando V_0 fixo temos:

$$\begin{aligned}\hat{\sigma}^2(\alpha) &= \frac{\sum_{i=1}^{N}(y_i - x_i'\hat{\beta})' V_0^{-1}(y_i - x_i'\hat{\beta})}{Nk} \\ &= \frac{SQR}{Nk}\end{aligned}$$

3. Substituindo em $l(\hat{\beta}, \alpha, \hat{\sigma}^2)$, obtemos:

$$\begin{aligned}l(\hat{\beta}, \alpha, \hat{\sigma}^2) &\propto \frac{-N}{2}\left(k \log\left(\frac{SQR}{kN}\right) + \log|V_0|\right) - \frac{kN}{2} \\ &\propto \frac{-N}{2}(k \log SQR - k\log(kN) + \log|V_0|) - \frac{kN}{2} \\ &\propto \frac{-N}{2}(k\log SQR + \log|V_0|)\end{aligned}$$

Verossimilhança Perfilada

O estimador de α envolve a maximização de

$$l(\hat{\beta}, \alpha, \hat{\sigma}^2)$$

- Finalmente após obtermos $\hat{V}_0$ encontramos β e σ^2.
- A maximização com respeito a α exige o cálculo de determinantes e inversas. Isto indica a necessidade de métodos numéricos.

Estruturação de V_0

1. **Simétrica Composta e AR(1)**: 1 componente:

$$\text{Total}=p+1+1$$

2. **Não Estruturada homocedástica**: $\frac{n(n-1)}{2}$ componentes:

$$\text{Total}=\frac{n(n-1)}{2}+p+1$$

3. **Não Estruturado heterocedástica** :

$$\text{Total}=p+\frac{n(n+1)}{2}$$

Propriedades de $\hat{\beta}_{EMV}$(ASSINTÓTICAS)

1. $\hat{\beta}_{EMV}$ é consistente para β;

2. $\hat{\beta}_{EMV}$ é assintóticamente normal (Wald):

$$\sqrt{Nk}(\hat{\beta}_{EMV} - \beta) \xrightarrow{\mathcal{D}} N(0, \Im^{-1})$$

3. Estas propriedades valem assintoticamente, mesmo se y_i não tiver distribuição normal multivariada (para dados completos).

4. Usamos as estatísticas:
WALD e da ***RV*** (Razão de Verossimilhança) para fazer inferência sobre β

Propriedades de $\hat{\beta}_{EMV}$(ASSINTÓTICAS)

5 As distribuições de referência (assintótica) normal e qui-quadrado são utilizadas como aproximações da t e da F, respectivamente. É possível estimar os gl para utilizar a t e a F, especialmente para amostras de tamanho pequeno.

6 O valor-p obtido através da estatística de Wald é menor do que o verdadeiro (e será tão menor quanto menor for o tamanho da amostra).

7 Devemos evitar o uso da estatística de Wald para testar os componentes de variância (α e σ^2) pois a convergência para normal é lenta para amostras pequenas e variâncias próximas de zero. Desta forma, o recomendado é a estatística da razão de verossimilhança.

Estimação Conjunta (β, α)

1. EMV;
2. EMVR (Estimador de Máxima Verossimilhança Restrita).

EMV (σ^2 é absorvido por V_0)

$$l(\beta,\alpha) = \frac{-kN}{2} log(2\pi) - \frac{N}{2} log|V_0| - \frac{1}{2}\sum_{i=1}^{N}(y_i - x_i\beta)' V_0^{-1}(y_i - x_i\beta)$$

Propriedades do EMV: $\hat{\beta}$ é consistente e assintoticamente Normal (para dados completos).

Estatísticas: Wald e RV (Inferência para β).

Estimador de Máxima Verossimilhança Restrita (ou Residual)

Estudo linear-normal transversal EMV (amostra de tamanho N)

$$\hat{\sigma}^2 = \frac{SQR}{N} = \frac{(y - x\hat{\beta})'(y - x\hat{\beta})}{N}$$

$$E[\hat{\sigma}^2] = \frac{N}{N-p}\sigma^2$$

- **Razão:** EMV não leva em consideração que β é estimado pelos dados;
- **Proposta:** utilizar EMVR (Estimador Máxima Verossimilhança Restrita);
- **Ideia:** Separar as partes dos dados para estimar α daqueles utilizados para estimar β.

Transformar a resposta Y tal que a distribuição resultante não dependa de β. Ou seja,

$$Z = AY \quad \text{tal que} \quad E(Z) = 0$$

Exemplo: Modelo Linear-Normal Transversal

$$A = I - H = I - X(X'X)^{-1}X'$$

$$E(Z) = (I - X(X'X)^{-1}X')E(Y) = X\beta - X(X'X)^{-1}X'X\beta = 0$$

Transformação $y \rightarrow (Z, \hat{\beta})$

A função de log-verossimilhança para Z escrita em termos de Y e $\hat{\beta}$ é:

$$l^*(\alpha) = \frac{1}{2}\sum_{i=1}^{N} log|V_0| - \frac{1}{2}\sum_{i=1}^{N}(y_i - x_i\hat{\beta})' V_0^{-1}(y_i - x_i\hat{\beta}) - \frac{1}{2}\left|\sum_{i=1}^{N} x_i{'} V_0^{-1} x_i\right|$$

Justificativa: Na ausência de informação de β, nenhuma informação sobre α é perdida se a inferência para α for feita por $l^*(\alpha)$ ao invés de $l(\alpha)$.

Obervação

O termo adicional da função de log-verossimilhança restrita:

$$-\frac{1}{2}\left|\sum_{i=1}^{N} x_i`V_0^{-1}x_i\right| = log|cov(\widehat{\beta})|^{1/2}.$$

Este termo é o equivalente a fazer a correção no denominador de $\widehat{\sigma}^2$.

Processo de Estimação: EMVR

1. Estimar β por:

$$
\begin{aligned}
\hat{\beta} &= (X'V^{-1}X)^{-1}XV^{-1}Y \\
&= \left(\sum_{i=1}^{N} x_i' V_0^{-1} x_i\right)^{-1} \left(\sum_{i=1}^{N} x_i' V_0^{-1} y_i\right);
\end{aligned}
$$

2. Encontra, o EMVR para α a partir de $l^*(\alpha)$;

3. Continuar este processo até a convergência.

Processo de Estimação: EMVR

1. O EMVR é recomendado para α quando comparado ao EMV. No entanto, a correção do vício se torna *desprezível* quando Nk é muito maior que p;

2. A estatística da Razão de MVR pode ser usado para comparar modelos de covariâncias aninhadas mas não pode ser utilizado para comparar modelos aninhados para a média. Neste caso devemos usar o EMV.

1. Características:
    - GEE e EMVR são similares (iqualmente eficientes) com dados completos.
    - A única condição para GEE produzir inferências válidas é a estrutura da média estar corretamente especificada.
    - Especificando corretamente a estrutura de variância-covariância ganha-se em eficiência no processo inferencial.
    - Na presença de dados faltantes (MAR e NMAR), o GEE não produz inferências válidas. Por outro lado, o EMVR produz inferências válidas nesta condição (somente MAR)se a distribuição normal for corretamente especificada para a resposta.

2 Limitações:

- Dados longitudinais desbalanceados. Somente a estrutura de variância-covariância

$$Cor(Y_{ij}, Y_{il}) = \rho^{|t_{ij} - t_{il}|}$$

  é possível ser especificada sob desbalanceamento. Disponível no pacote gls do R.

- Falta de flexibilidade na especificação da estrutura de variância-covariância da resposta.

Análise de Resíduos e Diagnóstico

Pontos Principais:

- A análise de dados longitudinais não fica completa sem a examinação dos resíduos. Ou seja, a verificação das suposições impostas ao modelo e ao processo de inferência.
- As ferramentas usuais de análise de resíduos para a regressão convencional (com observações independentes) podem ser estendidas para a estrutura longitudinal.

Suposições dos Modelos

- Estrutura da média: forma analítica, linearidade dos β's.
- Normalidade (resposta e efeitos aleatórios).
- Estrutura de Variância-Covariância: Homocedasticidade e correlação das medidas do mesmo indivíduo.

Suposições dos Modelos

- Defina o vetor de resíduos para cada indivíduo

$$r_i = Y_i - X_i\hat{\beta}, \quad i = 1, \ldots, N,$$

  que é um estimador para o vetor de erros

$$\epsilon_i = Y_i - X_i\beta, \quad i = 1, \ldots, N.$$

- Tratando-se de dados longitudinais, sabemos que os componentes do vetor de resíduos r_i são correlacionados e não necessariamente têm variância constante.

Utilidade dos Resíduos r_i

Gráficos:

- Gráfico de r_{ij} vs $\widehat{Y}_{ij}$: é útil para identificar alguma tendência sistemática (por exemplo, presença de curvatura) e presença de pontos extremos (”outliers”). O modelo corretamente especificado não deve apresentar nenhuma tendência neste gráfico.

- Limitação: este gráfico não tem necessariamente uma largura constante. Ou seja, cuidado ao interpretar este gráfico com relação a homocedasticidade.

- Gráfico de r_{ij} vs t_{ij}: é também útil para identificar violação da homocedasticidade.

Solução: Examinar resíduos transformados

- Há muitas possibilidades para transformar os resíduos.
- A tranformação deve ser realizada de forma que os resíduos "imitem"aqueles da regressão linear padrão.
- Os resíduos r_i^* definidos a seguir são não-correlacionados e têm variância unitária:
$$r_i^* = L_i^{-1} r_i,$$
em que L_i é a matriz triangular superior resultante da decomposição de Cholesky da matriz de covariâncias estimada $\widehat{Var}(Y_i)$, ou seja, $\widehat{Var}(Y_i) = L_i L_i'$.

Resíduos transformados

- Podemos aplicar a mesma transformação ao vetor de valores preditos $\hat{Y}_i$, ao vetor da variável resposta Y_i e à matriz de covariáveis $\mathbf{X}_i$:

$$\begin{aligned}
\hat{Y}_i^* &= L_i^{-1}\hat{Y}_i \\
Y_i^* &= L_i^{-1}Y_i \\
\mathbf{X}_i^* &= \hat{L}_i^{-1}\mathbf{X}_i
\end{aligned}$$

e então todos os diagnósticos de resíduos usuais para a regressão linear padrão podem ser aplicados para r_i^*.

Gráficos de Adequação

- Gráfico de dispersão dos resíduos transformados r_{ij}^* versus os valores preditos transformados $\hat{Y}_{ij}^*$: não deve apresentar nenhum padrão sistemático para um modelo corretamente especificado. Ou seja, deve apresentar um padrão aleatório em torno de uma média zero. Útil para verificar homocedasticidade.

- Gráfico de dispersão dos resíduos transformados r_{ij}^* versus covariáveis transformadas X_{ij}^* (em especial, idade ou tempo): verificar padrões de mudança na resposta média ao longo do tempo;

- QQ-plot de r_i^*: verificar normalidade e identificar outliers.

Semi-variograma

- O semi-variograma, denotado por $\gamma(h_{ijk})$, é dado por:

$$\gamma(h_{ijk}) = \frac{1}{2}E(r_{ij} - r_{ik})^2,$$

em que $h_{ijk} = t_{ij} - t_{ik}$.

- O semi-variograma pode ser utilizado como uma ferramenta para verificar a adequação do modelo selecionado para a estrutura de covariância dos dados.

Semi-variograma

- Como os resíduos têm média zero, o semi-variograma pode ser reescrito como:

$$\begin{aligned}\gamma(h_{ijk}) &= \frac{1}{2}E(r_{ij} - r_{ik})^2 \\ &= \frac{1}{2}E(r_{ij}^2 + r_{ik}^2 - 2r_{ij}r_{ik}) \\ &= \frac{1}{2}Var(r_{ij}) + \frac{1}{2}Var(r_{ik}) - Cov(r_{ij}, r_{ik}).\end{aligned}$$

- Quando o semivariograma é aplicado aos resíduos transformados, r_{ij}^*, a seguinte simplificação é obtida:

$$\gamma(h_{ijk}) = \frac{1}{2}(1) + \frac{1}{2}(1) - 0 = 1.$$

Semi-variograma

- Logo, se o modelo é corretamente especificado para a matriz de covariâncias, o gráfico do semi-variograma amostral $\hat{\gamma}(h_{ijk})$ dos resíduos transformados versus h_{ijk} deveria flutuar aleatoriamente em torno de uma linha horizontal centrada em 1.

- O semi-variograma é muito sensível a outliers.