Análise de Dados Longitudinais
Aula 17.10.2018

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

Sumário

Modelos Lineares Generalizados Longitudinais

1. Fácil transferência entre modelos (marginal e condicional) para resposta gaussiana.
2. Transferência díficil entre modelos quando a resposta não é gaussiana.
3. Modelos Marginais
   - Especificação completa: o ajuste por MV pode ser complicado.
   - Alternativa Não-Verossimilhança: MQG, GEE, etc.
4. Modelos Condicionais: ajuste complicado.

Resposta Longitudinal Não-Gaussiana

1. Equações de Estimação Generalizadas
2. Modelos Lineares Mistos Generalizados

Modelos Marginais: GEE

Equações de Estimação Generalizadas

$$\sum_{i=1}^{N} D_i' V_i (Y_i - \mu_i) = 0,$$

em que

- $D_i = \partial \mu_i / \partial \beta$ e $\mu_i = g^{-1}(X_i \beta)$, ou seja, o inverso da função de ligação g.
- $$Var(Y_i) = V_i = \phi A_i^{1/2}(\beta) R_i(\alpha) A_i^{1/2}(\beta)$$
  em que A_i é uma matriz diagonal formada por $Var(Y_{ij})$, R_i é matriz de correlação de trabalho e ϕ é um parâmetro de dispersão/escala.
- $Var(\widehat{\beta})$ é estimada pela variância robusta (estimador sanduíche).

Formas de Correlação de Trabalho

- *independência*, $\boldsymbol{R}_i(\boldsymbol{\alpha}) = \boldsymbol{I}_{n_i}$;
  $\Rightarrow$ dados longitudinais não correlacionados.
- *simetria composta*, especifica que $\boldsymbol{R}_i(\boldsymbol{\alpha}) = \rho 1_{n_i} 1_{n_i}' + (1-\rho) 1_{n_i}$;
  $\Rightarrow$ mesma correlação para qualquer par de tempo.
- *AR-1*, para a qual $\boldsymbol{R}_i(\boldsymbol{\alpha}) = \rho^{|j-j'|}$;
  $\Rightarrow$ válida para medidas igualmente espaçadas no tempo;
- *não estruturada*;
  $\Rightarrow$ estima todas as $n_i(n_i-1)/2$ correlações de $\boldsymbol{R}$.

Variância do Estimador

1. *Naive* ou "baseada no modelo"
   $$\widehat{Var}(\hat{\beta}) = \left(\sum_{i=1}^{N} \hat{\boldsymbol{D}}_i' \boldsymbol{R}_i(\hat{\alpha})^{-1} \hat{\boldsymbol{D}}_i \right)^{-1}. \tag{1}$$
2. *Robusta* ou "empírica"
   $$\widehat{Var}(\hat{\beta}) = \boldsymbol{M}_0^{-1} \boldsymbol{M}_1 \boldsymbol{M}_0^{-1}, \tag{2}$$

em que

$$\begin{aligned} \boldsymbol{M}_0 &= \sum_{i=1}^{N} \hat{\boldsymbol{D}}_i' \boldsymbol{R}_i(\hat{\alpha})^{-1} \hat{\boldsymbol{D}}_i, \\ \boldsymbol{M}_1 &= \sum_{i=1}^{N} \hat{\boldsymbol{D}}_i' \boldsymbol{R}_i(\hat{\alpha})^{-1} (\boldsymbol{y}_i - \hat{\boldsymbol{\mu}}_i)(\boldsymbol{y}_i - \hat{\boldsymbol{\mu}}_i)' \boldsymbol{R}_i(\hat{\alpha})^{-1} \hat{\boldsymbol{D}}_i. \end{aligned}$$

Exemplo: Bernoulli-logit

1. $\mu_{ij} = E(Y_{ij}) = P(Y_{ij} = 1)$.
2. $$logit(\mu_{ij}) = log(\mu_{ij}/(1-\mu_{ij})) = X_{ij}'\beta$$
   $$\mu_{ij} = \frac{e^{X_{ij}'\beta}}{1 + e^{X_{ij}'\beta}}$$
3. $$Var(Y_{ij}) = \mu_{ij}(1-\mu_{ij})$$
4. $$\nu_{ij} = \mu_{ij}(1-\mu_{ij}) \qquad A_i = diag(\nu_{i1}, \nu_{i2} \ldots, \nu_{in})$$

Exemplo: Poisson-log

1. $log(\mu_{ij}) = X'_{ij}\beta$.
$$\mu_{ij} = e^{X'_{ij}\beta}$$
2. $$Var(Y_{ij}) = \mu_{ij} = e^{X'_{ij}\beta}$$
3. $$\nu_{ij} = e^{X'_{ij}\beta} = \mu_{ij}$$

Estimando a Correlação de Trabalho

- Liang e Zeger (1986) utilizaram estimativas de momentos para os parâmetros da matriz de correlação de trabalho.
- Ou seja, utilizar estimadores baseados nos resíduos para as quantidades envolvidas em R_i.
- Resíduos de Pearson:
$$e_{ij} = \frac{y_{ij} - \hat{\mu}_{ij}}{\sqrt{\hat{\nu}_{ij}}},$$
em que $\nu_{ij} = \mu_{ij}(1 - \mu_{ij})$ para resposta binária e $\nu_{ij} = \mu_{ij}$, para contagem.

Estimadores de Momentos usando Resíduos

- **Independência**: $Cor(Y_{ij}, Y_{ij'}) = 0$.
- **Simetria Composta:** $Cor(Y_{ij}, Y_{ij'}) = \alpha$.
$$\hat{\alpha} = \frac{\sum_{i=1}^{N} \sum_{j>j'} e_{ij} e_{ij'}}{\sum_{i=1}^{N} n_i(n_i - 1)/2 - p}.$$
- **Não estruturada:** $Cor(Y_{ij}, Y_{ij'}) = \alpha_{jj'}$.
$$\hat{\alpha}_{jj'} = \frac{\sum_{i=1}^{N} e_{ij} e_{ij'}}{N - p}.$$

Podemos estimar ϕ por
$$\hat{\phi} = \sum_{i=1}^{N} \sum_{j=1}^{n_i} e_{ij}^2/(N - p).$$
em que $N = \sum_{i=1}^{N} n_i$ e p é a dimensão de β.

Ajustando GEE

1. Use MLE para encontrar a estimativa inicial para β (assumindo independência)
2. Encontre os resíduos e estime α e ϕ.
3. Atualize a estimativa de β.
4. Faça iterações em (2)-(3) até a convergência.
5. Estime $Var(\hat{\beta})$ usando o estimador sanduíche.

Modelos Lineares Generalizados Mistos

1. Modelos Lineares Generalizados
   - Resposta na família exponencial: normal, gama, exponencial, Bernoulli, Poisson, etc.
   - Preditor Linear: $X_i'\beta$.
   - Função de Ligação: $g(\mu_i) = X_i'\beta$.
2. Modelos Lineares Generalizados Mistos

   Preditor Linear:

$$X_i\beta + Z_i b_i.$$

Modelos Generalizado Misto Longitudinal

1.

$$g(E(Y_{ij}|b_i)) = X_{ij}'\beta + Z_{ij}'b_i$$

em que:

$(\beta)_{p\times 1}$: efeitos fixos;

$(b_i)_{q\times 1}$: efeitos aletaórios.

e,

2.

$$b_i \sim N_q(0, \Sigma) \text{ e } \varepsilon_{ij} \sim N(0, \sigma^2)$$

Sendo b_i e ε_{ij} independentes.

Função de Verossimilhança

$$\begin{aligned}
L(\theta|y) &= \prod_{i=1}^{N} p(y_i|\theta) \\
&= \prod_{i=1}^{N} \int p(y_i, b_i|\theta)db_i \\
&= \prod_{i=1}^{N} \int p(y_i|b_i, \theta)p(b_i|\theta)db_i
\end{aligned}$$

em que,

$$p(y_i|b_i, \theta) \sim \text{Bernoulli-logit/Poisson-log, etc}$$

e

$$p(b_i|\theta) \sim N_q(0, \Sigma)$$

Solução

- No modelo linear-normal, a integral pode ser resolvida analiticamente.
- Em geral, aproximações são necessárias no caso não-normal.
- Aproximação do integrando: Laplace
- Aproximação da integral: quadratura gaussiana.

Usualmente, a combinação normal-logit não tem solução simples.

Interpretação dos Parâmetros

- O vetor β no GEE tem interpretação populacional. Ou seja, a mesma interpretação dos modelos transversais.
- O vetor β no modelo GLMM tem interpretação condicional sob o nível dos efeitos aleatórios. Ou seja, interpretação específica para cada indivíduo.
- Portanto, as estimativas dos modelos são diferentes!

A seguir, aprofundamos as conexões entre os dois modelos.

Conexões entre modelos de efeitos aleatórios e modelos marginais

No modelo misto modelamos a média condicional, $\mu_{ij} = E(Y_{ij}|b_i)$.

Invertendo a função de ligação, obtemos

$$E(Y_{ij}|b_i) = g^{-1}(X_{ij}'\beta + Z_{ij}'b_i).$$

Marginalmente, ponderando sobre os efeitos aleatórios, a média é

$$E(Y_{ij}) = E\left[E(Y_{ij}|b_i)\right] = \int g^{-1}(X_{ij}'\beta + Z_{ij}'b_i)f(b_i, \Sigma)db_i,$$

em que $f(b_i, \Sigma)$ é a $N_q(0, \Sigma)$, densidade dos efeitos aleatórios.

Conexões entre modelos de efeitos aleatórios e modelos marginais

Para a função de ligação identidade,

$$E(Y_{ij}) = \int (X_{ij}'\beta + Z_{ij}'b_i)f(b_i, \Sigma)db_i = X_{ij}'\beta.$$

O modelo marginal tem a mesma forma e efeitos β. Isto não é verdade para outras ligações.

Por exemplo, para o modelo logístico

$$E(Y_{ij}) = E\left[\frac{\exp(X_{ij}'\beta + Z_{ij}'b_i)}{1 + \exp(X_{ij}'\beta + Z_{ij}'b_i)}\right],$$

Esta esperança não tem a forma $\exp(X_{ij})/\left[1 + \exp(X_{ij})\right]$, exceto quando b_i tem uma distribuição degenerada ($\sigma_b = 0$).

Conexões entre modelos de efeitos aleatórios e modelos marginais

Zeger et. al. (1988) mostraram que, para o modelo condicional,

$$logito(\mu_{ij}) \approx a(\Sigma)X_{ij}'\beta,$$

em que $a(\Sigma) = |c^2\Sigma Z_{ij}Z_{ij}' + I|^{-q/2}$ e $c = \frac{16\sqrt{3}}{15\pi}$.

No caso particular em que $q = 1$ (intercepto aleatório), temos a relação aproximada:

$$\beta_M \approx \frac{\beta_{EA}}{\sqrt{1 + \frac{16\sqrt{3}}{15\pi}\sigma_b^2}}.$$

Conexões entre modelos de efeitos aleatórios e modelos marginais

Assim, se $Var(b_i) = \sigma_b > 0$, o efeito marginal β_M é menor que o efeito condicional β_{EA}.

Além disso, a discrepância entre β_{EA} e β_M aumenta quando σ_b cresce.

Por exemplo, se $\sigma_b^2 = 3,5$, então $\beta_{EA} \approx 1,49\beta_M$; se $\sigma_b^2 = 9$, então $\beta_{EA} \approx 2,03\beta_M$.

A figura a seguir ilustra por que o efeito marginal é menor que o efeito condicional.

Modelo Logístico: Efeito Condicional e Marginal