CE075 - Análise de Dados Longitudinais

Silva, J.L.P.

28 de agosto, 2019

Modelos de Regressão

Notação

Seja Y_{ij} a variável resposta para i-ésimo indivíduo $(i = 1, \ldots, N)$ na j-ésima ocasião $(j = 1, \ldots, n_i)$.

Dado que temos n_i medidas repetidas da resposta no mesmo indivíduo, podemos agrupá-las em um vetor $n_i \times 1$, denotado por

$$\boldsymbol{Y}_i = \begin{pmatrix} Y_{i1} \\ Y_{i2} \\ \vdots \\ Y_{in_i} \end{pmatrix},$$

ou, por conveniência,

$$\boldsymbol{Y}_i = (Y_{i1}, Y_{i2}, \ldots, Y_{in_i})'.$$

Dependência e Correlação

O principal interesse está na média da resposta (em particular, em mudanças da média no tempo e como esta mudança depende de covariáveis).

Denote a média ou esperança de cada resposta Y_{ij} por $\mu_j = E(Y_{ij})$.

Adicionalmente, para permitir que a resposta média varie de indivíduo para indivíduo como função de covariáveis medidas em nível de indivíduo, requeremos $\mu_{ij} = E(Y_{ij})$.

Denotando a média condicional de Y_{ij} por μ_{ij}, a variância condicional de Y_{ij} é definida como:

$$\sigma_j^2 = E\left[Y_{ij} - E(Y_{ij})\right]^2 = E\left(Y_{ij} - \mu_{ij}\right)^2.$$

Dependência e Correlação

A *covariância* condicional entre as respostas em duas ocasiões diferentes, digamos Y_{ij} e Y_{ik}, é definida por

$$\sigma_{jk} = E\left[(Y_{ij} - \mu_{ij})(Y_{ik} - \mu_{ik})\right].$$

e fornece uma medida da dependência *linear* entre Y_{ij} e Y_{ik}, dado as covariáveis.

A correlação condicional entre Y_{ij} e Y_{ik} é denotada por

$$\rho_{jk} = \frac{E\left[(Y_{ij} - \mu_{ij})(Y_{ik} - \mu_{ik})\right]}{\sigma_j \sigma_k},$$

que, por definição, assume valores entre -1 e $+1$.

Dependência e Correlação

Definimos a matriz de variância-covariância como segue

$$
\begin{aligned}
Cov\begin{pmatrix} Y_{i1} \\ Y_{i2} \\ \vdots \\ Y_{in_i} \end{pmatrix} &= \begin{pmatrix} Var(Y_{i1}) & Cov(Y_{i1}, Y_{i2}) & \dots & Cov(Y_{i1}, Y_{in_i}) \\ Cov(Y_{i2}, Y_{i1}) & Var(Y_{i2}) & \dots & Cov(Y_{i2}, Y_{in_i}) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(Y_{in_i}, Y_{i1}) & Cov(Y_{in_i}, Y_{i2}) & \dots & Var(Y_{in_i}) \end{pmatrix} \\
&= \begin{pmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1n_i} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2n_i} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n_i 1} & \sigma_{n_i 2} & \dots & \sigma_{n_i n_i} \end{pmatrix}.
\end{aligned}
$$

Assumimos que as variâncias e covariâncias são constantes. Note que há simetria, ou seja, $Cov(Y_{ij}, Y_{ik}) = \sigma_{jk} = \sigma_{kj} = Cov(Y_{ik}, Y_{ij})$.

Dependência e Correlação

Usaremos frequentemente a notação

$$Cov(\boldsymbol{Y}_i) = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1n_i} \\ \sigma_{21} & \sigma_2^2 & \dots & \sigma_{2n_i} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n_i1} & \sigma_{n_i2} & \dots & \sigma_{n_i}^2 \end{pmatrix}.$$

Definimos a matriz de correlação em termos de

$$Corr(\boldsymbol{Y}_i) = \begin{pmatrix} 1 & \rho_{12} & \dots & \rho_{1n_i} \\ \rho_{21} & 1 & \dots & \rho_{2n_i} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n_i1} & \rho_{n_i2} & \dots & 1 \end{pmatrix},$$

que é simétrica, ou seja, $Corr(Y_{ij}, Y_{ik}) = \rho_{jk} = \rho_{kj} = Corr(Y_{ik}, Y_{ij})$.

Modelo Marginal

O modelo para a resposta média em cada ocasião não incorpora a dependência sobre nenhum efeito aleatório ou sobre respostas anteriores.

Apropriado quando o foco da análise é inferir sobre a população média.

1. $E(Y_{ij}|X_{ij}) = \mu_{ij}$, com $g(\mu_{ij}) = \eta_{ij} = X'_{ij}\beta$.
2. $Var(Y_{ij}|X_{ij}) = \phi v(\mu_{ij})$, em que ϕ é um parâmetro de dispersão e $v(\cdot)$ é uma função conhecida da média.
3. A correlação intra-indivíduos é função de α. Por exemplo:
    - $Corr(Y_{ij}, Y_{ik}) = \alpha^{|k-j|}$ (AR-1 para respostas contínuas);
    - $logOR(Y_{ij}, Y_{ik}) = \alpha_{jk}$ (não-estruturado para respostas categóricas).

Modelo Marginal

A caracterização de um modelo marginal envolve:

1. Modelar a resposta média $E(\boldsymbol{Y}_i)$.
2. Modelar a estrutura de Variância-Covariância $Var(\boldsymbol{Y}_i), \ i = 1, \ldots N$.
3. Assumir uma distribuição (normal) para a resposta (dispensável).

Dois caminhos:

1. Assumir resposta normal: usar MQG ou MV (usual ou restrita).
2. Não assumir distribuição para a resposta: usar GEE: "Generalized Estimation Equations".

Modelos Mistos

Incluem efeitos aleatórios no modelo de efeitos fixos, em nível de indivíduo, modelando a heterogeneidade entre indivíduos e induzindo, assim, uma estrutura de covariância entre as respostas repetidas.

1. $E(Y_{ij}|X_{ij}, b_i) = \mu_{ij}$, com $g(\mu_{ij}) = \eta_{ij} = X'_{ij}\beta + Z'_{ij}b_i$, sendo b_i o efeito aleatório associado com Y_i.
2. $Var(Y_{ij}|X_{ij}, b_i) = \phi v(\mu_{ij})$.
3. Geralmente assume-se $b_i \sim N_q(0, G)$.

Modelos de Transição

A distribuição condicional de Y_{ij} é descrita como uma função explícita das respostas passadas e de um vetor de variáveis preditoras.

1. $E(Y_{ij}|X_{ij}, H_{ij}) = \mu_{ij}$, com $H_{ij} = \{Y_{id}, d = 1, \ldots, j-1\}$.
   $g(\mu) = X_{ij}'\beta + \sum_{q=1}^{Q} f_q(H_{ij}, \alpha)$, em que $f_q(\cdot)$ são funções conhecidas.
2. $Var(Y_{ij}|X_{ij}, H_{ij}) = \phi v(\mu_{ij})$.
3. A correlação entre $Y_{i1}, \ldots, Y_{in_i}$ é avaliada através do parâmetro α que aparece na função $f_q(\cdot)$.