

Equações de Estimação para Dados Longitudinais

Pedro Luciano de Oliveira Gomes

Centro de Ciências Exatas e da Terra
Universidade Federal do Rio Grande do Norte - UFRN
Departamento de Estatística
Seminário de Tópicos Especiais - STE

25 de agosto de 2016

Sumário

1 Introdução

2 Dados Longitudinais

- Equações de Estimação Independentes
- Equações de Estimação Generalizadas
- Conclusão

3 Referências Bibliográficas

Sumário

1 Introdução

2 Dados Longitudinais

- Equações de Estimação Independentes
- Equações de Estimação Generalizadas
- Conclusão

3 Referências Bibliográficas

Introdução

Os modelos lineares generalizados (MLG) são propostos para análise de dados com apenas uma observação por indivíduo os quais assumem independência entre as observações.

Introdução

Quando tratamos de dados longitudinais a variável resposta, y_{it} , e um vetor $p \times 1$ de covariáveis, x_{it} , são observados em $t = 1, \dots, n$ tempos para cada sujeito $i = 1, \dots, K$.

Introdução

Por exemplo:

- Crescimento de uma planta a cada dia;
- Acompanhamento nutricional de uma criança mensalmente.

Introdução

Dessa forma, devemos levar em consideração a correlação entre os valores observados para um determinado sujeito para podermos estimar os parâmetros do modelo.

Introdução

O objetivo deste trabalho é apresentar uma extensão dos modelos lineares generalizados para a distribuição marginal de y_{it} . A distribuição conjunta das medidas repetidas não é especificada. Em vez disso, será introduzido equações de estimação que nos dão estimadores consistentes para os parâmetros da regressão e suas respectivas variâncias, sob suposição fraca da distribuição conjunta.

Sumário

1 Introdução

2 Dados Longitudinais

- Equações de Estimação Independentes
- Equações de Estimação Generalizadas
- Conclusão

3 Referências Bibliográficas

Dados Longitudinais

Para estabelecer notação, seja $y_i = (y_{i1}, y_{i2}, \dots, y_{it})'$ o vetor de dimensões $n_i \times 1$ de respostas e $x_i = (x_{i1}, x_{i2}, \dots, x_{it})'$ a matriz de dimensões $n_i \times p$ de covariáveis para o i -ésimo indivíduo ($i = 1, \dots, n$) para a j -ésima observação (ou tempo), $j = 1, \dots, t$. Assume que a densidade marginal de y_{it} é dada por:

$$f(y_{it}) = \exp[\{y_{it}\theta_{it} - a(\theta_{it}) + b(y_{it})\}\phi].$$

em que $\theta_{it} = h(x_{it}\beta)$. Por esta fórmula os dois momentos de y_{it} são dados por

$$E(y_{it}) = a'(\theta_{it}), \quad \text{var}(y_{it}) = a''(\theta_{it})/\phi$$

Equações de Estimação Generalizadas

As equações de estimação podem levar em conta a correlação para melhorar a eficiência dos estimadores. Vale ressaltar que os estimadores de β permanecem consistentes.

Dados Longitudinais

- Equações de estimação independentes (EEI);
- Equações de estimação generalizadas (EEG).

Equações de Estimação Independentes (EEI)

O estimador $\hat{\beta}_I$ de β tem como hipótese de que as observações repetidas de um sujeito são independentes umas das outras.

Equações de Estimação Independentes (EEI)

Sob hipótese de independência, a função score da verossimilhança analisada é dada por

$$\begin{aligned}U_i(\beta) &= \sum_{i=1}^k X_i^T \Delta_i (Y_i - a'_i(\theta)) \\ &= \sum_{i=1}^k X_i^T \Delta_i (Y_i - \mu_i) = 0\end{aligned}$$

em que $\Delta_i = \text{diag}(d\theta_{it}/d\eta_{it})$, $\eta_{it} = x_{it}\beta$, é uma matrix $n \times n$.

Equações de Estimação Independentes (EEI)

O estimador $\hat{\beta}_I$ é definido como a solução dessa equação. Sob leves condições de regularidade, temos o teorema a seguir.

Equações de Estimação Independentes (EEI)

Teorema 1

O estimador $\hat{\beta}_I$ de β é consistente e $\sqrt{N}(\hat{\beta}_I - \beta)$ é assintoticamente normal multivariada com média 0 e matriz de covariância V_I , conforme $N \rightarrow \infty$, dada por

$$V_I = \lim_{N \rightarrow \infty} N \left(\sum_{i=1}^N X_i^T \Delta_i A_i \Delta_i X_i \right)^{-1} \left(\sum_{i=1}^N X_i^T \Delta_i \text{cov}(Y_i) \Delta_i X_i \right) \left(\sum_{i=1}^N X_i^T \Delta_i A_i \Delta_i X_i \right)^{-1}$$

Equações de Estimação Independentes

No entanto, a principal desvantagem de $\hat{\beta}_I$ é que este pode não ser tão eficiente em casos que a autocorrelação é alta. Para isto, é proposto as equações de estimação generalizadas.

Equações de Estimação Generalizadas

As Equações de Estimação Generalizadas (EEG) foram introduzidas por Liang & Zeger (1986), em que apenas o primeiro e o segundo momento da função densidade marginal precisam ser definidos.

Equações de Estimação Generalizadas

Seja $R(\alpha)$ uma matriz simétrica de dimensões $n_i \times n_i$ que satisfaz as condições de uma matriz de correlação e, seja α um vetor de dimensões $n_i \times 1$ que caracteriza $R(\alpha)$.

Também conhecida como matriz correlação de “trabalho”.

Equações de Estimação Generalizadas

Define-se

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

que será igual a $Cov(Y_i)$, se $R(\alpha)$ é, de fato, a matriz de correlação verdadeira para os Y_{it} 's. Tem-se que ϕ é o parâmetro de dispersão ou escala e $A_i = \text{diag}\{a''(\theta_{it})\}$.

Equações de Estimação Generalizadas

Definem-se as equações de estimação generalizadas como

$$\sum_{i=1}^n D_i^T V_i^{-1} (Y_{it} - \mu_{it}) = 0,$$

em que $D_i = d\{a_i'(\theta)\}/d\beta = A_i \triangleq X_i$.

Observa-se que a equação reduz-se às equações de independência se $R(\alpha)$ for a matriz de identidade.

Equações de Estimação Generalizadas

A expressão dada pode ser reescrita em função de β . Substituindo α nas equações por $\hat{\alpha}(Y, \beta, \phi)$, um estimador $\sqrt{N}$ - consistente para α quando β e ϕ são conhecidos. Consequentemente, a equação

$$\sum_{i=1}^n D_i^T V_i^{-1} (Y_{it} - \mu_{it}) = 0,$$

tem a forma

$$\sum_{i=1}^n U_i \left[\beta, \hat{\alpha}\{\hat{\beta}, \hat{\phi}(\beta)\} \right] = 0,$$

e $\hat{\beta}_G$, as estimativas dos parâmetros para dados correlacionados, é a solução da equação.

Apesar do objetivo principal da utilização das EEGs ser a análise de dados correlacionais com mais de uma observação por indivíduo, estas podem também ser usadas para modelar a superdispersão.

O uso destas equações tornou-se frequente e os programas computacionais estatísticos trazem sua implementação devido aos estimadores robustos das variâncias dos parâmetros serem consistentes, mesmo se a matriz de correlação utilizada for mal especificada.

Sumário

1 Introdução

2 Dados Longitudinais

- Equações de Estimação Independentes
- Equações de Estimação Generalizadas
- Conclusão

3 Referências Bibliográficas

Referências

-  LIANG, Kung-ye; ZEGGER, Scott L.. **Longitudinal data analysis using generalized linear models**. Biometrika, Great Britain, v. 1, n. 73, p.13-22, jan. 1986.
-  Costa, Silvano Cesar da. **Modelos Lineares Generalizados Mistos Para Dados Longitudinais**. Piracicaba - São Paulo, jan. 2003.