Análise de Dados Longitudinais
Aula 27.08.2018

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

Sumário

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 p-dimensional de parâmetros.

O EMV (Estimador de Máxima Verossimilhança) é aquele $\hat{\theta}$ 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

A função escore é dada por

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

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

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

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

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

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

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

Revisão - Teoria de Verossimilhança: Medida de Incerteza

Usualmente é difícil encontrar o valor esperado na distribuição de Y. No entanto, podemos utilizar qualquer estimador consistente de I.

Usamos a matriz de informação observada

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

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

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

Obs. O resultado é verdadeiro para qualquer estimador consistente de I.

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:

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

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.

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

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

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

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

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)

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

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.

Estimador de Máxima Verossimilhança Restrita (ou Residual)

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 $l(\alpha)$.

Observação

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. Encontrar 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 *Nn* 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.

Modelos Marginais: GEE/EMVR

1. GEE e EMVR são similares (igualmente eficientes) com dados completos.
2. A única condição para GEE produzir inferências válidas é a estrutura da média estar corretamente especificada.
3. Especificando corretamente a estrutura de variância-covariância ganha-se em eficiência no processo inferencial.
4. 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.