CE225 - Modelos Lineares Generalizados

Cesar Augusto Taconeli

31 de agosto, 2018

Aula 7 - Inferência em modelos lineares generalizados

Testes de hipóteses

- Vamos discutir neste momento testes para hipóteses do tipo:

$$H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0 \quad H_1 : \boldsymbol{\beta} \neq \boldsymbol{\beta}_0 \tag{1}$$

- Nas hipóteses apresentadas, $\boldsymbol{\beta}$ representa um ou mais parâmetros do modelo ajustado, e $\boldsymbol{\beta}_0$ valores postulados (fixados) para esses parâmetros na hipótese nula.

Testes de hipóteses

- Apenas para ilustração, considere o seguinte modelo:

$$ln(\mu) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 \quad (2)$$

- Exemplos de hipóteses:

$$H_0 : \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \quad vs \quad H_1 : \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{pmatrix} \neq \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix};$$

Testes de hipóteses

$$H_0 : \begin{pmatrix} \beta_1 \\ \beta_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad vs \quad H_1 : \begin{pmatrix} \beta_1 \\ \beta_3 \end{pmatrix} \neq \begin{pmatrix} 0 \\ 0 \end{pmatrix};$$

$$H_0 : \beta_2 = 0 \quad vs \quad H_1 : \beta_2 \neq 0;$$

$$H_0 : \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \\ 0 \\ 2 \end{pmatrix} \quad vs \quad H_1 : \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{pmatrix} \neq \begin{pmatrix} 1 \\ -1 \\ 0 \\ 2 \end{pmatrix} \ldots$$

- Os principais testes de hipóteses em modelos lineares generalizados são:
  - Teste da razão de verossimilhanças;
  - Teste de Wald;
  - Teste escore.

Testes da razão de verossimilhanças

- Seja L_0 a verossimilhança maximizada sob a hipótese nula, e L_1 a verossimilhança maximizada de forma não restrita (permitindo que H_0 ou H_1 seja verdade).
- A razão $\Lambda = L_0/L_1 \leq 1$, uma vez que L_0 resulta da maximização sobre um conjunto restrito de valores para $\boldsymbol{\beta}$.

Testes da razão de verossimilhanças

- A estatística do teste da razão de verossimilhança é definida por:

$$-2 ln \Lambda = -2 ln(L_0/L_1) = -2(l_0 - l_1), \qquad (3)$$

sendo l_0 e l_1 as log-verossimilhanças.

- Sob H_0 e ϕ conhecido, a estatística do teste tem distribuição assintótica ($n \to \infty$) χ^2 com q graus de liberdade, sendo q o número de parâmetros fixados em H_0.

Nota: se o parâmetro de dispersão é desconhecido (sendo estimado), o teste da razão de verossimilhança tem melhor aproximação pela distribuição F.

Testes de Wald

- Seja $\beta = (\beta_0, \beta_1)$, com β_0 e β_1 compondo uma partição do vetor de parâmetros original β, com q e $p - q$ parâmetros, respectivamente.
- Para testar $H_0 : \beta_\mathbf{0} = \mathbf{0}$, a estatística do teste de Wald fica definida por

$$\hat{\beta}_0{}' \widehat{Var}^{-1}(\hat{\beta}_\mathbf{0})\hat{\beta}_0, \tag{4}$$

em que $\hat{\beta}_\mathbf{0}$ é a estimativa de β_0 produzida pela maximização da verossimilhança irrestrita e $\widehat{Var}(\hat{\beta}_\mathbf{0})$ é o bloco da matriz de variâncias, correspondente aos elementos de $\hat{\beta}_\mathbf{0}$. também baseada na verossimilhança irrestrita.

- A estatística de Wald, sob H_0 e considerando ϕ conhecido, também tem distribuição assintótica ($n \to \infty$) χ^2 com q graus de liberdade, sendo q o número de parâmetros fixados em H_0.

Testes de Wald

- Para o teste de um único parâmetro, com hipótese nula $H_0 : \beta_k = \beta_0$, a estatística do teste de Wald fica dada por:

$$z = \frac{\hat{\beta}_k - \beta_0}{\sqrt{\widehat{Var}(\hat{\beta}_k)}} \tag{5}$$

tendo, sob H_0, distribuição assintótica $N(0, 1)$ quando ϕ é conhecido.

Testes escore

- A estatística do teste escore é definido por:

$$\boldsymbol{S}'(\hat{\boldsymbol{\beta}}_{\mathbf{0}})\widehat{Var}_0(\hat{\beta})\boldsymbol{S}(\hat{\boldsymbol{\beta}}_{\mathbf{0}}), \tag{6}$$

em que $\boldsymbol{S}(\hat{\boldsymbol{\beta}}_{\mathbf{0}})$ e $\widehat{Var}_0(\hat{\beta})$ são a função escore e a matriz de variâncias avaliadas **sob o modelo restrito** (sob H_0).

- O teste escore não requer o ajuste do modelo sob H_1, sendo conveniente quando H_1 define modelos bem mais complexos do que H_0.
- O teste escore sob H_0 também tem distribuição assintótica $(n \rightarrow \infty)$ χ^2 com q graus de liberdade, sendo q o número de parâmetros fixados em H_0.

Ilustração - Testes

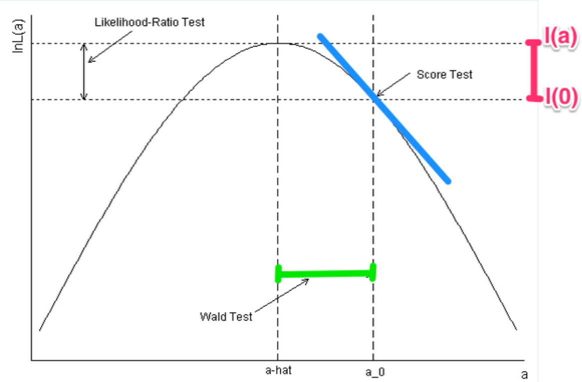

Figura 1: Log-verossimilhança e informação usada nos três testes para a hipótese $H_0 : \beta = a_0$.

Intervalos de confiança

- Intervalos de confiança para qualquer dos três métodos podem ser obtidos invertendo as respectivas estatísticas de teste.
- Por exemplo, um intervalo de confiança 95% para um único parâmetro β_k, é definido pelo conjunto de valores β_0 tais que $H_0 : \beta_k = \beta_0$ não é rejeitada ao nível de significância de 5%.
- Um intervalo de confiança assintótico $100(1-\alpha)\%$ para β_k, **baseado no teste de Wald**, tem limites:

$$IC(\beta_k; 1-\alpha) = \hat{\beta}_k \pm z_{\alpha/2}\sqrt{\widehat{Var}(\hat{\beta}_k)}, \tag{7}$$

em que $z_{\alpha/2}$ é o quantil $\alpha/2$ da distribuição normal padrão.

Intervalos de confiança

- Pode-se obter um intervalo de confiança para β_k baseado na **verossimilhança perfilada**.
- Seja $H_0 : \beta_k = \beta_0$ e Ψ representando o conjunto dos demais parâmetros do modelo.
- Ao inverter o teste da razão de verossimilhanças, para determinar o conjunto de valores β_0 que compõem o intervalo de confiança, a estimativa de máxima verossimilhança de Ψ varia para os diferentes valores de β_0.
- O intervalo de confiança baseado na verossimilhança perfilada de β_k é definido pelo conjunto de valores β_0 tais que:

$$-2[L(\beta_0, \hat{\Psi}(\beta_0)) - L(\hat{\beta}_k, \hat{\Psi})] < \chi^2_1(\alpha), \tag{8}$$

sendo $L(\beta_0, \hat{\Psi}(\beta_0))$ a verossimilhança maximizada para $\beta_k = \beta_0$ e $L(\hat{\beta}_k, \hat{\Psi})$ a verossimilhança maximizada de forma irrestrita.

Intervalos de confiança

- Além de intervalos de confiança para os parâmetros, é interessante também obter intervalos de confiança para $\mu_{\boldsymbol{x}} = E[y|\boldsymbol{x}]$, sendo $\boldsymbol{x} = (x_1, x_2, ..., x_p)'$ um específico vetor de covariáveis.
- A estimativa pontual para $\mu_{\boldsymbol{x}}$ é dada por:

$$\hat{\mu}_{\boldsymbol{x}} = g^{-1}(\boldsymbol{x'}\hat{\boldsymbol{\beta}}). \tag{9}$$

- Seja $\hat{\eta}_{\boldsymbol{x}} = \boldsymbol{x'}\hat{\boldsymbol{\beta}}$. Como $\hat{\eta}_{\boldsymbol{x}}$ é uma combinação linear dos $\hat{\beta}'s$, decorre que, assintoticamente:

$$\hat{\eta}_{\boldsymbol{x}} \sim Normal(\boldsymbol{x'}\boldsymbol{\beta}, \boldsymbol{x'}Var(\hat{\boldsymbol{\beta}})\boldsymbol{x}) \tag{10}$$

Intervalos de confiança

- Um intervalo de confiança assintótico $100(1-\alpha)\%$ para $\eta_{\boldsymbol{x}} = \boldsymbol{x}'\beta$ fica dado por:

$$IC(\eta_{\boldsymbol{x}}; 1-\alpha) = \boldsymbol{x}'\hat{\beta} \pm z_{\alpha/2}\sqrt{\boldsymbol{x}'\widehat{Var}(\hat{\beta})\boldsymbol{x}}. \tag{11}$$

- Dessa forma, um intervalo de confiança assintótico $100(1-\alpha)\%$ para $\mu_{\boldsymbol{x}} = g^{-1}(\eta_{\boldsymbol{x}})$ fica dado por:

$$IC(\mu_{\boldsymbol{x}}; 1-\alpha) = (g^{-1}(LI); g^{-1}(LS)), \tag{12}$$

se $g(\cdot)$ é uma função estritamente crescente, e:

$$IC(\mu_{\boldsymbol{x}}; 1-\alpha) = (g^{-1}(LS); g^{-1}(LI)), \tag{13}$$

se $g(\cdot)$ é uma função estritamente decrescente, onde LI e LS denotam os limites de confiança inferior e superior para $\eta_{\boldsymbol{x}}$.