Mestrado Acadêmico em Bioestatística

Estatísticas da Razão de Verossimilhança, Wald e Escore em amostras pequenas para a distribuição Beta

Wesley Oliveira Furriel
André Felipe B. Menezes
Rosangela Getirana Santana

DEPARTAMENTO DE ESTATÍSTICA
UNIVERSIDADE ESTADUAL DE MARINGÁ – UEM

7 de Dezembro de 2016

Organização

- Nas mais variadas áreas do conhecimento o emprego de indicadores, proporções ou taxas tem sido amplamente utilizado para a sumarização, identificação e hierarquização de distintos fenômenos analisados pelos pesquisadores.
- Exemplos desses casos são o IDH (Índice de Desenvolvimento Humano) Cobertura Vacinal, Taxa de Mortalidade e Coeficiente de Gini.
- Para a modelagem destes casos, a família de distribuições Beta é apresentada como uma das alternativas mais adequadas.

Uma variável aleatória X que tem distribuição Beta tem função densidade de probabilidade indexada por dois parâmetros dada por (CASELLA e BERGER, 2002):

$$f(x \mid \alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} \quad (1)$$

sendo $0 < x < 1$ e α e β ambos parâmetros de forma positivos. A esperança e variância de X são respectivamente dadas por:

$$\mathbb{E}(X) = \frac{\alpha}{(\alpha+\beta)} \text{ e } \mathrm{Var}(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}. \quad (2)$$

Ferrari e Cribari-Neto (2004) propõe uma reparametrização da distribuição Beta (1), considerando-se a média da resposta e um parâmetro de dispersão. Sejam $\mu = \alpha/(\alpha+\beta)$ e $\phi = \alpha + \beta$, ou seja, $\alpha = \mu\,\phi$ e $\beta = (1-\mu)\,\phi$. Desse modo, a função densidade de probabilidade de Y pode ser escrita da forma:

$$f(y \mid \mu, \phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} y^{\mu\phi-1}(1-y)^{(1-\mu)\phi-1}, \; 0 < y < 1 \qquad (3)$$

com $0 < \mu < 1$ e $\phi > 0$. A esperança e variância de Y são dadas respectivamente por:

$$\mathbb{E}(Y) = \mu \quad \text{e} \quad \mathrm{Var}(Y) = \frac{\mathrm{V}(\mu)}{(1+\phi)} \qquad (4)$$

em que $\mathrm{V}(\mu) = \mu(1-\mu)$, de modo que μ é a média da variável resposta e ϕ o parâmetro de dispersão.

Objetivos

- Iniciar a discussão acerca do uso da distribuição Beta para a modelagem de indicadores sociais e da saúde;
- Avaliar o Erro Tipo I dos testes via simulação Monte Carlo, conforme:
  - Tamanho amostral (n);
  - Parâmetros μ e ϕ.
- Identificar o teste com o melhor desempenho conforme os cenários (n, μ, ϕ).

Função de Verossimilhança

Considere $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)$ uma amostra aleatória de n observações provenientes de (3), têm-se que o lnaritmo natural da função verossimilhança pode ser escrito na forma:

$$
\begin{aligned}
l(\mu, \phi \mid \boldsymbol{y}) = & \, n \ln \Gamma(\phi) - n \ln \Gamma(\mu\phi) - n \ln \Gamma(\phi - \mu \phi) \\
& + (\mu \phi - 1) \sum_{i=1}^{n} \ln \, y_i + [\phi - \mu \phi - 1] \sum_{i=1}^{n} \ln (1 - y_i)
\end{aligned} \tag{5}
$$

Função de Escore

Diferenciando (5) em relação a μ e ϕ temos as componentes do vetor escore:

$$\mathcal{U}_{\mu}(\boldsymbol{\theta} \mid \boldsymbol{y}) = n\,\phi\,\Psi\,(\phi - \mu\,\phi) - n\,\phi\,\Psi\,(\mu\,\phi) + \phi \sum_{i=1}^{n} \ln\, y_i - \phi \sum_{i=1}^{n} \ln\,(1 - y_i)$$

$$\mathcal{U}_{\phi}(\boldsymbol{\theta} \mid \boldsymbol{y}) = n\,\Psi\,(\phi) + n\,(\mu - 1)\,\Psi(\phi - \mu\,\phi) - n\,\mu\,\Psi(\mu\,\phi) + \mu \sum_{i=1}^{n} \ln\, y_i$$

$$+ (1 - \mu) \sum_{i=1}^{n} \ln\,(1 - y_i).$$

em que $\Psi(\cdot)$ é a função digama. Os estimadores de máxima verossimilhança de $\boldsymbol{\theta} = (\mu, \phi)$, denotados por $\widehat{\boldsymbol{\theta}} = (\widehat{\mu}, \widehat{\phi})$, são obtidos igualando o vetor escore a zero e resolvendo simultaneamente o sistema de equações não lineares.

Matriz de Informação

Além disso, a matriz de informação observada de $\boldsymbol{\theta}$ através da amostra $\boldsymbol{y}$, definida como sendo o negativo da segunda derivada da função log-verossimilhança, é dada por:

$$\mathcal{I}(\boldsymbol{\theta} \mid \boldsymbol{y}) = \begin{bmatrix} K_{\mu\,\mu} & K_{\mu\,\phi} \\ K_{\phi\,\mu} & K_{\phi\,\phi} \end{bmatrix} \tag{6}$$

em que:

- $K_{\mu\,\mu} = n\,\phi^2\,\Psi^{'}\,(\mu\,\phi) + n\,\phi^2\,\Psi^{'}\,(\phi - \mu\,\phi)$.
- $K_{\mu\,\phi} = n\,\mu\,\phi\,\Psi^{'}(\mu\,\phi) + n\,\Psi(\mu\,\phi) - n\,(1-\mu)\,\phi\,\Psi^{'}(\phi - \mu\,\phi) - n\,\Psi(\phi - \mu\,\phi) - \sum_{i=1}^{n} \ln\, y_i + \sum_{i=1}^{n} \ln\,(1 - y_i) = K_{\phi\,\mu}$.
- $K_{\phi\,\phi} = n\,\mu^2\,\Psi^{'}\,(\mu\,\phi) + n\,(\mu - 1)^2\,\Psi^{'}\,(\phi - \mu\,\phi) - n\,\Psi^{'}\,(\phi)$.

sendo $\Psi^{'}(\cdot)$ a função trigama.

Testes de Hipóteses

Seja $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)$ uma amostra aleatória de n observações independentes proveniente da densidade Beta reparametrizada. Nosso interesse é testar:

$H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$

$H_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$

Em que $\boldsymbol{\theta} = (\mu, \phi)$ e $\boldsymbol{\theta}_0 = (\mu_0, \phi_0)$

Para tal utilizamos os testes da Razão de verossimilhança, Wald e Escore.

O teste da razão de verossimilhança avalia o máximo das funções ln-verossimilhanças dos modelos restrito e completo, a estatística é dada por:

$$TRV = 2\left[l(\widehat{\boldsymbol{\theta}} \mid \boldsymbol{y}) - l(\boldsymbol{\theta}_0 \mid \boldsymbol{y})\right] \tag{7}$$

Sob a hipótese nula, TRV tem distribuição assintótica qui-quadrado com $p = 2$ graus de liberdade (PAWITAN, 2001).

O teste de Wald consiste em comparar a estimativa do parâmetro $(\widehat{\boldsymbol{\theta}})$ com o valor do parâmetro sob a hipótese nula $(\boldsymbol{\theta}_0)$. A estatística do teste tem forma quadrática definida por:

$$W = \left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)^T \mathcal{I}(\widehat{\boldsymbol{\theta}}) \left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) \tag{8}$$

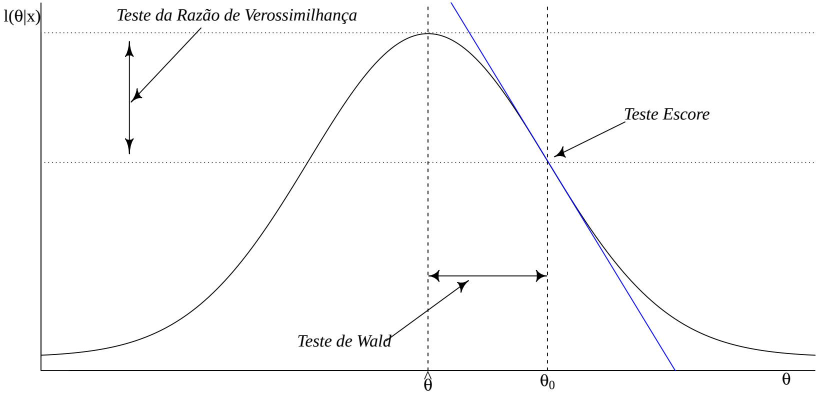
sendo $\mathcal{I}(\widehat{\boldsymbol{\theta}})$ a matriz de informação localmente nas estimativas de máxima verossimilhança $\widehat{\mu}$ e $\widehat{\phi}$. Desse modo, sob a hipótese nula, W tem assintoticamente distribuição qui-quadrado com $p = 2$ graus de liberdade (MILLAR, 2011).

Por fim, o teste Escore baseia-se na conjectura da função escore $\mathcal{U}(\boldsymbol{\theta}_0)$ e da matriz de informação esperada $\mathcal{I}(\widehat{\boldsymbol{\theta}_0})$ sob a hipótese nula. Sendo a estatística do teste:

$$S = \left(\mathcal{U}(\boldsymbol{\theta}_0)\right)^T \mathcal{I}(\boldsymbol{\theta}_0)^{-1} \mathcal{U}(\boldsymbol{\theta}_0) \tag{9}$$

em que, $\mathcal{U}(\boldsymbol{\theta}_0)$ é a função escore e $\mathcal{I}(\boldsymbol{\theta}_0)^{-1}$ é a inversa da matriz de informação sob a hipótese nula. A estatística S, sob a hipótese nula, tem assintoticamente distribuição qui-quadrado com $p = 2$ graus de liberdade (MILLAR, 2011).

Testes de Hipótese



Foi conduzido um estudo de simulação Monte Carlo que averiguou o comportamento do tamanho do teste (*Erro Tipo I*) em diferentes cenários.

Amostras de tamanhos:
$n = 20, 40, 60, 80$ e 100

Parâmetros:
$\mu = 0.05, 0.25, 0.5, 0.75$ e 0.95
$\phi = 5, 15$ e 50

Para cada combinação (n, μ, ϕ) foram geradas $B = 10000$ amostras pseudoaleatórias da nova parametrização da distribuição Beta. Em todas as instâncias adotamos níveis de significância nominais $\alpha = 0,05$.

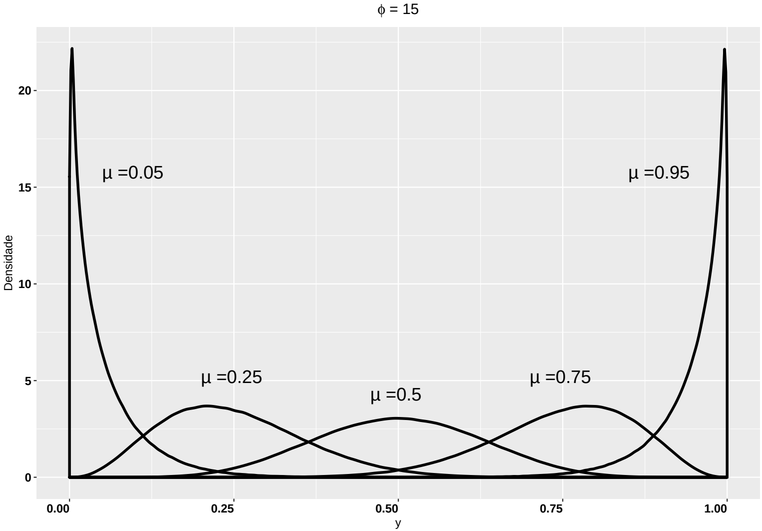
Podemos estimar o *Erro Tipo I* (α) gerando amostras independentes, sob a hipótese nula, e calculando a proporção de vezes que H_0 foi rejeitada erradamente, isto é,

$$\widehat{\alpha} = \frac{\text{Número de vezes em que } H_0 \text{ é rejeitada } \mid H_0 \text{ é verdadeira}}{B} \tag{10}$$

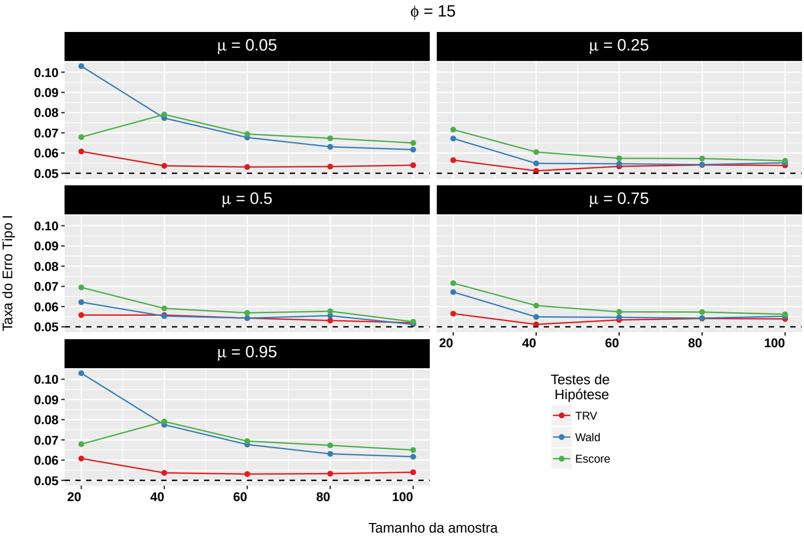
```
%macro simulacao_beta;
%do mu = 0.05 , 0.25,0.5 ,0.75 ,0.95;
      %do phi = 5 , 15, 50;
            %do n=20 %to 100 %by 20;
                  data simulacao;
                  do b=1 to 10000;
                   do i=1 to &n;
                    y = valores aleatórios da Beta;
                  end;end;
                  run;
             proc nlmixed data=simulacao;
             -Teste TRV;
             -Teste Wald;
             by b;
             run;
                 proc iml;
                 -Teste Escore;
                 quit;
                  data resultados;
                  -Junção dos resultados dos testes;
                  run;
                   proc freq data=resultados;
                   -Taxa do Erro Tipo I;
                   run;
            %end;
      %end;
%end;
%mend;
```

Comportamento

Densidade da Beta para $(\mu, \phi = 15)$

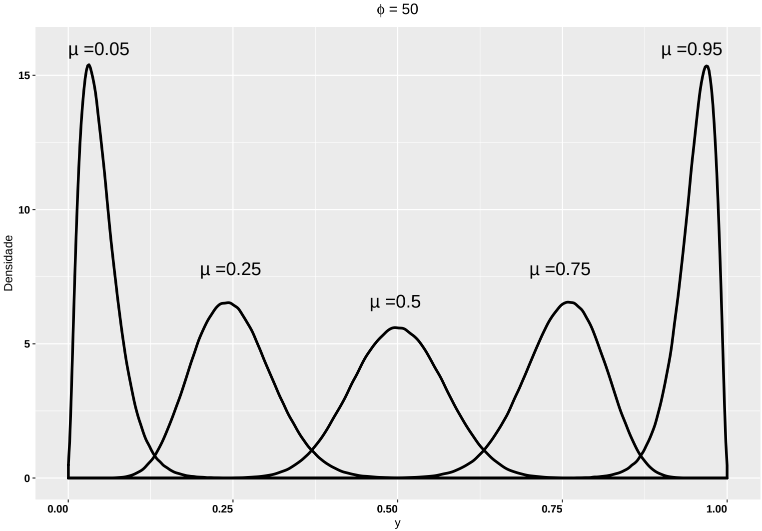


Resultados

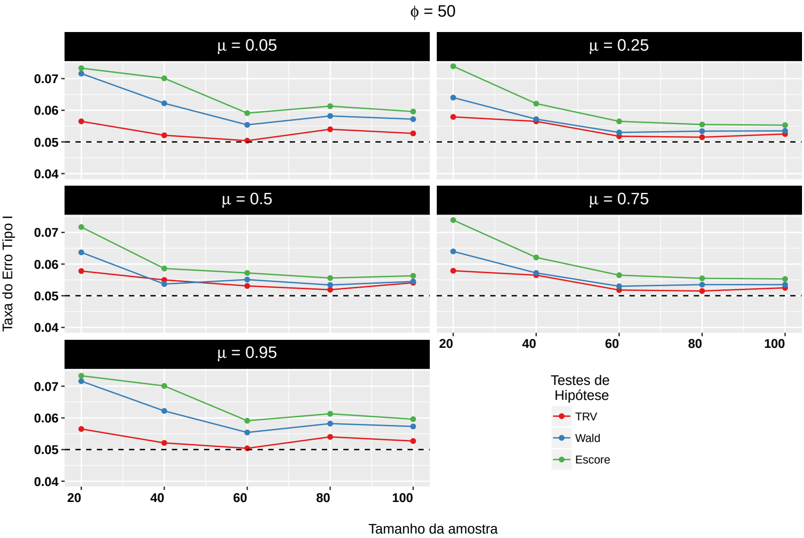


ϕ = 15
μ = 0.05
μ = 0.25
μ = 0.5
μ = 0.75
μ = 0.95
0.10
0.09
0.08
0.07
0.06
0.05
20
40
60
80
100
Taxa do Erro Tipo I
Tamanho da amostra
Testes de Hipótese
TRV
Wald
Escore

Densidade da Beta para $(\mu, \phi = 50)$



Resultados


ϕ = 50
μ = 0.05
μ = 0.25
μ = 0.5
μ = 0.75
μ = 0.95
0.07
0.06
0.05
0.04
20
40
60
80
100
Taxa do Erro Tipo I
Tamanho da amostra
Testes de Hipótese
TRV
Wald
Escore

Conclusões

- De um modo geral, o teste da Razão de Verosimilhança apresentou melhor desempenho;
- Verificamos que em amostras pequenas os testes apresentam resultados mais distantes do nível de significância nominal, $\alpha = 0,05$;
- No cenário, no qual, o parâmetro de dispersão é $\phi = 5$, o teste Escore não se aproximou do nível de significância nominal;

Referências

[1] CASELLA, G.; BERGER, R. L. Statistical inference. Pacific Grove, CA: Duxbury, 2002.

[2] FERRARI, S.; CRIBARI-NETO, F. Beta regression for modelling rates and proportions. Journal of Applied Statistics, v. 31, n. 7, p. 799-815, 2004.

[3] MAZUCHELI, J.; LOUZADA, F. Discriminação entre as distribuições Odd Weibull e Weibull. Rev. Bras. Biom 32.2 (2014): 226-237.

[4] MILLAR, R. B. Maximum likelihood estimation and inference: with examples in R, SAS and ADMB. Vol. 111. John Wiley & Sons, 2011.

[5] PAWITAN, Y. In all likelihood: statistical modelling and inference using likelihood. Oxford University Press, 2001.

Muito obrigado!