

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

Wesley Oliveira Furriel
André Felipe Menezes
Rosângela Getirana Santana

DEPARTAMENTO DE ESTATÍSTICA
UNIVERSIDADE ESTADUAL DE MARINGÁ – UEM

7 de Dezembro de 2016



Introdução

Contextualização

Família de Distribuições Beta

Distribuição Beta Reparametrizada

Objetivos

Função de Verossimilhança

Função Escore

Matriz de Informação

Testes de Hipóteses

Teste da Razão de Verossimilhança

Teste de Wald

Teste Escore

Estudo de Simulação

Cenários

Implementação

Resultados

Introdução

Contextualização



2

22

- 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.

Introdução

Família de Distribuições Beta



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

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

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

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

Introdução

Distribuição Beta Reparametrizada



4

22

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:

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

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

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

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



- ⚡ 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 da Razão de Verossimilhança, Wald e Escore 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) .



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

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



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

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

$$\begin{aligned} \mathcal{U}_{\phi}(\boldsymbol{\theta} \mid \mathbf{y}) &= n\Psi(\phi) + n(\mu - 1)\Psi(\phi - \mu\phi) - n\mu\Psi(\mu\phi) + \mu \sum_{i=1}^n \ln y_i \\ &\quad + (1 - \mu) \sum_{i=1}^n \ln(1 - y_i). \end{aligned}$$

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



Além disso, a matriz de informação observada de θ através da amostra y , definida como sendo o negativo da segunda derivada da função log-verossimilhança, é dada por:

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

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.



Seja $\mathbf{y} = (y_1, y_2, \dots, y_n)$ 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.

Testes de Hipóteses

Teste da Razão de Verossimilhança

10



22

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

$$TRV = 2 \left[l(\hat{\theta} | y) - l(\theta_0 | y) \right] \quad (7)$$

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

Testes de Hipótese

Teste de Wald

11



22

O teste de Wald consiste em comparar a estimativa do parâmetro ($\hat{\theta}$) com o valor do parâmetro sob a hipótese nula (θ_0). A estatística do teste tem forma quadrática definida por:

$$W = \left(\hat{\theta} - \theta_0 \right)^T \mathcal{I}(\hat{\theta}) \left(\hat{\theta} - \theta_0 \right) \quad (8)$$

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

Testes de Hipótese

Teste Escore



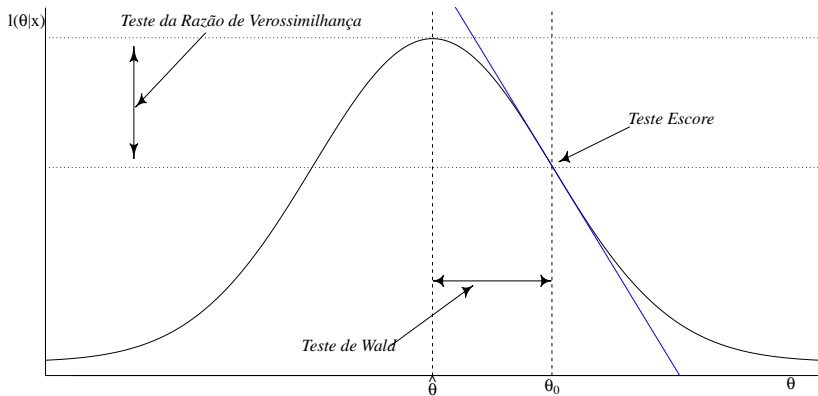
12

22

Por fim, o teste Escore baseia-se na conjectura da função escore $\mathcal{U}(\theta_0)$ e da matriz de informação esperada $\mathcal{I}(\hat{\theta}_0)$ sob a hipótese nula. Sendo a estatística do teste:

$$S = (\mathcal{U}(\theta_0))^T \mathcal{I}(\theta_0)^{-1} \mathcal{U}(\theta_0) \quad (9)$$

em que, $\mathcal{U}(\theta_0)$ é a função escore e $\mathcal{I}(\theta_0)^{-1}$ 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).





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 $\alpha = 0,05$.

Estudo de Simulação

Taxa do Erro Tipo I



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,

$$\hat{\alpha} = \frac{\text{Número de vezes em que } H_0 \text{ é rejeitada} \mid H_0 \text{ é verdadeira}}{B} \quad (10)$$

Estudo de Simulação

Implementação



22

16

```
%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;
```

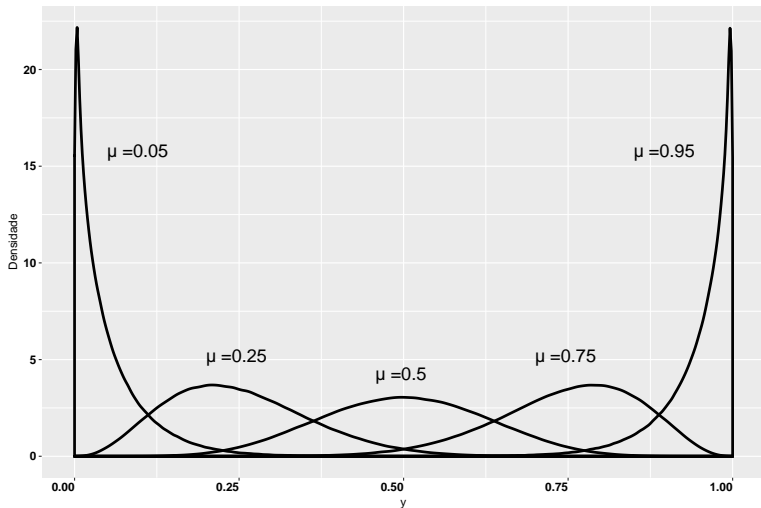
Comportamento

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



22

17

 $\phi = 15$ 

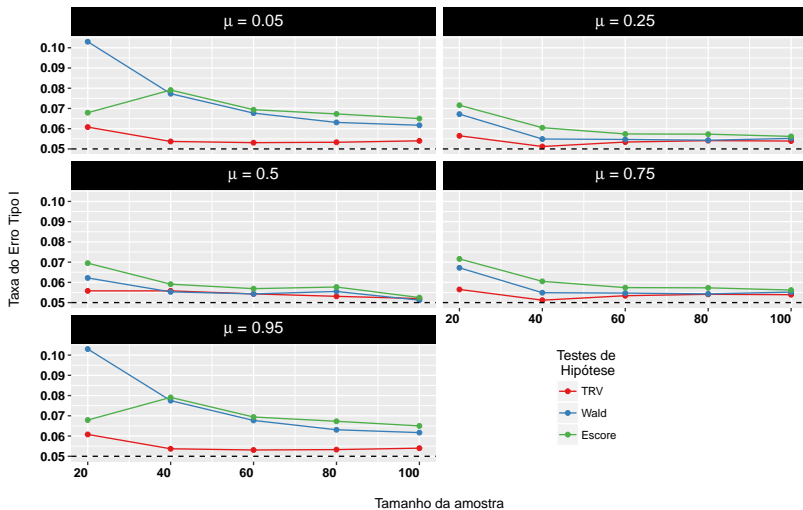
Resultados



22

18

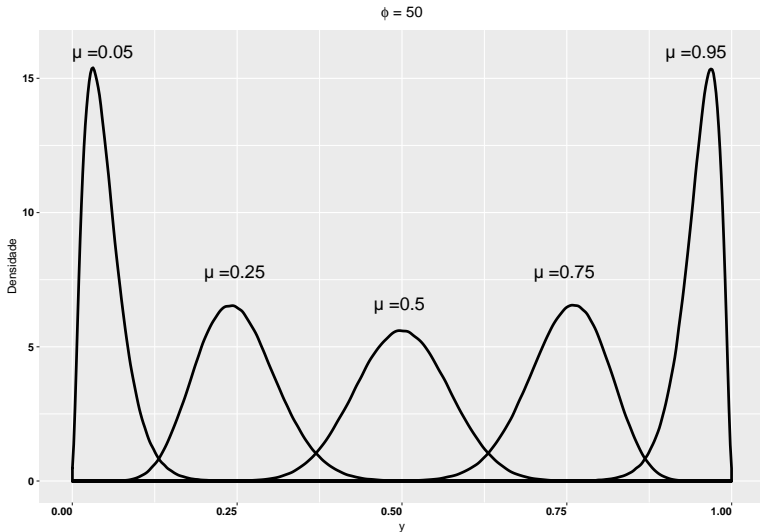
$$\phi = 15$$



Tamanho da amostra

Comportamento

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

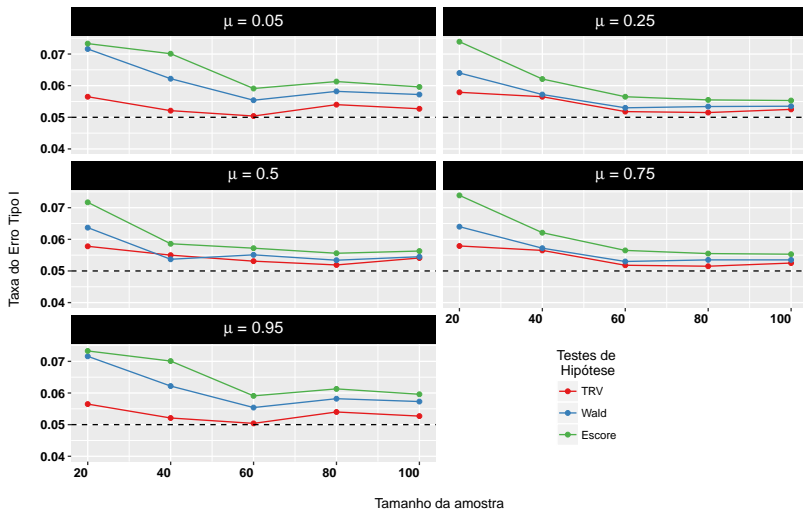


Resultados



22

20

 $\phi = 50$ 



- De um modo geral, o teste da Razão de Verossimilhança 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;



- [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!

