

# Testes de hipóteses para os parâmetros da distribuição Beta

English title here

## Resumo

*Sabendo-se a importância dos indicadores sociais e da saúde para compreensão dos fenômenos que cercam a sociedade contemporânea e a flexibilidade da distribuição de probabilidade Beta para sua modelagem. Este trabalho investigou as estatísticas dos testes da Razão de Verossimilhança, Wald, Escore e Gradiente, analisando a probabilidade do Erro Tipo I ( $\alpha$ ), com base na condução de um estudo de simulação Monte Carlo, visando avaliá-los em distintos cenários, variando o tamanho das amostras ( $n$ ) e os valores dos parâmetros  $\mu$  e  $\phi$ . Para a execução desse procedimento adotou-se uma reparametrização da Beta proposta por Ferrari e Cribari-Neto (2004) presente na análise de regressão. Os resultados obtidos revelaram diferenças entre os testes em relação ao  $\alpha$ , quando o tamanho da amostra é pequeno, bem como melhores resultados no panorama geral, para o teste da Razão de Verossimilhança.*

**Palavras-chave:** Distribuição Beta, simulação Monte Carlo, teste da Razão de Verossimilhança, teste de Wald, teste Escore, teste Gradiente.

## Abstract

*Knowing the importance of health indicators to understand the phenomena that are in contemporary society and the flexibility of Beta distribution to model them. This study investigated how statistics Likelihood Ratio tests, Wald, score and Gradient, analyzing the probability of Type I Error based on the conduction of a Monte Carlo simulation study, trying to verify them in different scenarios, varying the size of the samples ( $n$ ) and the value of the parameters  $\mu$  e  $\phi$ . For the execution of this study we did a reparametrization of the Beta, proposed by Ferrari and Cribari-Neto (2004) present in the regression analysis. The results showed differences between the tests compared to  $\alpha$  when the sample size is small, as well as better results in the general scenario, for the Gradient test.*

**Keywords:** Word1, word2, word3, word4, word5.

# 1 Introdução

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. Isso ocorre devido ao fato de que, variáveis que assumem um intervalo definido entre um valor mínimo e um máximo, permitem diagnósticos rápidos a cerca de determinado evento, uma vez que, elas podem, por exemplo, indicar o quão adequado ou inadequado determinado caso está na medida em que ele se aproxima dos valores presentes em umas destas extremidades.

Para melhor compreender essa ideia podemos citar alguns exemplos de variáveis que assumem essas características, como: o IDH (Índice de Desenvolvimento Humano), Cobertura Vacinal, Taxa de Mortalidade, Coeficiente de Gini entre outros.

Dentre as diversas distribuições de probabilidade disponíveis na literatura, a mais conhecida e também mais utilizada para modelagem de variáveis restrita a um intervalo é a família de distribuições Beta. Algumas aplicações da distribuição Beta são discutidas por exemplo em Johnson et al. (1995), Bury (1999) e Gupta e Nadarajah (2004). Uma variável aleatória  $X$  que pertencente a família de distribuições Beta indexada por dois parâmetros tem função densidade de probabilidade definida por:

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

sendo  $0 < x < 1$ ,  $\alpha$  e  $\beta$  ambos parâmetros de forma positivos e  $\Gamma(\alpha) = \int_0^\infty u^{\alpha-1} e^{-u} du$  a função gamma completa. A esperança e variância de  $X$  são respectivamente dadas por:

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

Apesar do intervalo da distribuição ser definido entre 0 e 1, é possível realizar uma generalização dessa característica para um  $Y$  que esteja restrito a um intervalo finito  $(m, n)$ , nos permitindo assim, modelar um conjunto de valores que estejam restritos a um intervalo identificável.

Ferrari e Cribari-Neto (2004) propõe uma reparametrização da distribuição Beta (1) para sua utilização no modelo de regressão. Já que, na análise de regressão, geralmente é mais útil modelar resposta média, então, a fim de obter uma estrutura de regressão para tal, juntamente com um parâmetro de dispersão, é interessante trabalhar com uma parametrização diferente da densidade Beta original. O modelo segue as propriedades da Beta, sendo então, adequado para casos em que a variável resposta  $Y$  é medida continuamente no intervalo  $0 < Y < 1$ . Seguindo a nova parametrização,  $\mu = \alpha/(\alpha + \beta)$  e  $\phi = \alpha + \beta$ , ou seja,  $\alpha = \mu\phi$  e  $\beta = (1 - \mu)\phi$  a média e a variância de  $Y$  são dadas pelas seguintes expressões,

$$\mathbb{E}(Y) = \mu \quad \text{e} \quad \text{Var}(Y) = \frac{V(\mu)}{(1 + \phi)} \quad (3)$$

de modo que  $\mu$  é a média da variável resposta e  $\phi$  pode ser interpretado como um parâmetro de dispersão. Por outro lado, a função densidade de probabilidade de  $Y$  é escrita na forma:

$$f(y | \mu, \phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} y^{\mu\phi-1} (1-y)^{(1-\mu)\phi-1}, \quad 0 < y < 1 \quad (4)$$

com  $0 < \mu < 1$  e  $\phi > 0$ .

Assim como na parametrização original é possível obter diferentes formas no comportamento da distribuição de acordo com os valores de seus parâmetros  $\mu$  e  $\phi$ . Ferrari e Cribari-Neto (2004) apontam que a distribuição pode ser simétrica quando  $\mu = 1/2$  e assimétrica quando  $\mu \neq 1/2$ . Ressalta-se, também que para  $\mu = 1/2$  e  $\phi = 2$  temos a distribuição Uniforme padrão. Além disso, a dispersão da distribuição, para um  $\mu$  fixado diminui quando os valores de  $\phi$  aumentam. Podemos visualizar esse comportamento na figura 1 em que foram fixados os valores de  $\phi$  e variados os de  $\mu$  para cada um dos gráficos.

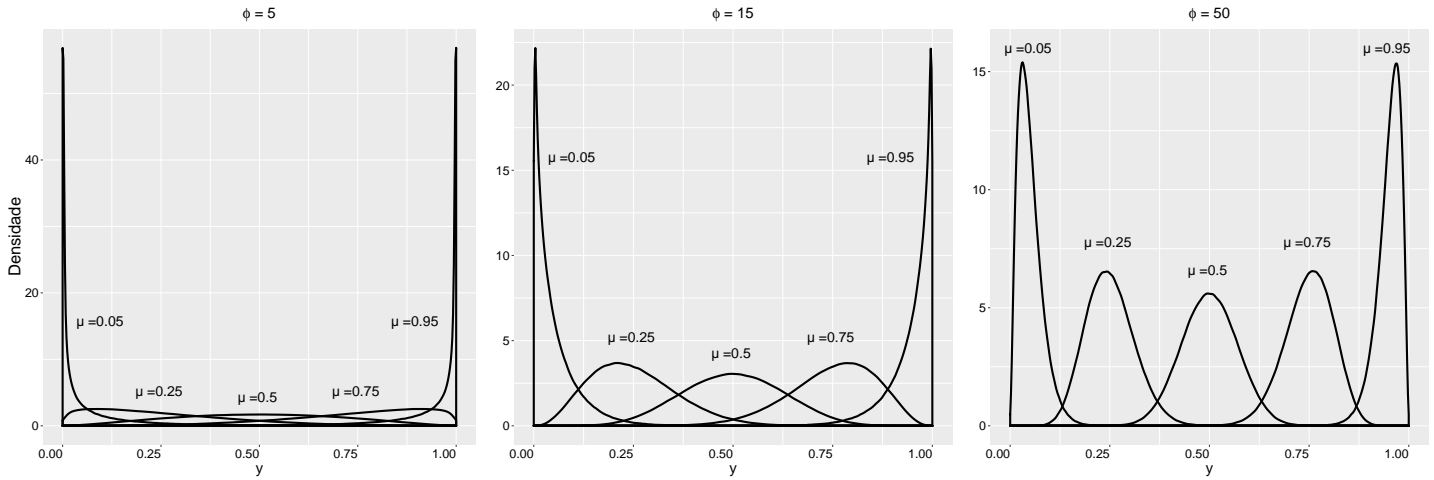


Figura 1: Comportamento da função densidade de probabilidade da distribuição Beta para diferentes valores de  $\mu$  e  $\phi$

Tendo em vista o panorama apresentado e a grande importância da distribuição Beta no contexto da Estatística e modelagem de dados, o presente trabalho teve como objetivo avaliar os resultados dos testes de hipótese de Razão de Verossimilhança, Wald, Escore e Gradiente, a partir da condução de um estudo de simulação Monte Carlo, que nos permitiu avaliar os testes, em distintos cenários, nos quais foram variados o tamanho das amostras ( $n$ ) e os valores dos parâmetros  $\mu$  e  $\phi$ .

Na literatura existem trabalhos semelhantes para outras distribuições. Podemos mencionar: distribuição Gamma (Bhaumik et al., 2009), distribuição Weibull Estendida de Marshall Olkin (Caroni, 2010), generalizações da distribuição Weibull (Krit et al., 2013), distribuição Odd Weibull (Mazucheli e Louzada, 2014) e distribuição Lindley Ponderada (Mazucheli et al., 2016).

Para a melhor sistematização e organização da discussão e dos resultados, o trabalho foi dividido em quatro seções. Na seção 2 apresenta-se a função de verossimilhança da distribuição Beta, as componentes do vetor escore e a matriz de informação observada utilizados nos testes avaliados. Os testes da razão de verossimilhança, Wald, Escore e Gradiente são descritos na seção 3. Na seção 4 nos ocupamos em detalhar os cenários sob os quais foi realizado o estudo de simulação bem como os resultados obtidos. Já na Seção 5 foi realizada uma aplicação com os teste em um conjuntos de dados reais. Por fim, apresentamos as considerações finais.

## 2 A função de verossimilhança

Considere  $\mathbf{y} = (y_1, y_2, \dots, y_n)$  uma amostra aleatória de  $n$  observações provenientes da distribuição Beta com vetor de parâmetro  $\boldsymbol{\theta} = (\mu, \phi)$  e função densidade de probabilidade apresentada em (4), então o logaritmo natural da função verossimilhança pode ser escrito na forma:

$$l(\boldsymbol{\theta} | \mathbf{y}) = n \log \Gamma(\phi) - n \log \Gamma(\mu\phi) - n \log \Gamma(\phi - \mu\phi) + (\mu\phi - 1) \sum_{i=1}^n \log y_i + [\phi - \mu\phi - 1] \sum_{i=1}^n \log(1 - y_i). \quad (5)$$

Diferenciando (5) em relação a  $\mu$  e  $\phi$  temos o vetor Escore  $\mathbf{U}(\boldsymbol{\theta} | \mathbf{y}) = (U_\mu, U_\phi)$  com componentes:

$$U_\mu = \frac{\partial}{\partial \mu} l(\boldsymbol{\theta} | \mathbf{y}) = n\phi\psi(\phi - \mu\phi) - n\phi\psi(\mu\phi) + \phi \sum_{i=1}^n \log y_i - \phi \sum_{i=1}^n \log(1 - y_i) \quad (6)$$

$$U_\phi = \frac{\partial}{\partial \phi} l(\boldsymbol{\theta} | \mathbf{y}) = n\psi(\phi) + n(\mu - 1)\psi(\phi - \mu\phi) - n\mu\psi(\mu\phi) + \mu \sum_{i=1}^n \log y_i + (1 - \mu) \sum_{i=1}^n \log(1 - y_i) \quad (7)$$

em que  $\psi(\cdot)$  representa a função digama, isto é,  $\psi(w) = d \log \Gamma(w) / dw$ .

Os estimadores de máxima verossimilhança (EMVs) de  $\boldsymbol{\theta} = (\mu, \phi)$ , denotados por  $\hat{\boldsymbol{\theta}} = (\hat{\mu}, \hat{\phi})$ , podem ser obtidos resolvendo simultaneamente as equações  $U_\mu = 0$  e  $U_\phi = 0$  para  $\mu$  e  $\phi$ . Estas equações não podem ser resolvidas de forma analítica, portanto procede-se com algum algoritmo de otimização não linear, como por exemplo Newton-Raphson.

A normalidade assintótica dos EMV de  $\boldsymbol{\theta}$  pode ser utilizada para construção de intervalos de confiança e testes de hipóteses sobre os parâmetros  $\mu$  e  $\phi$ . Sob condições de regularidade temos que  $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \overset{a}{\rightsquigarrow} N_2(0, \mathbf{I}^{-1}(\boldsymbol{\theta}))$ , em que  $\overset{a}{\rightsquigarrow}$  significa aproximadamente distribuído e  $\mathbf{I}(\boldsymbol{\theta})$  é a matriz de informação esperada. O comportamento assintótico permanece válido se

$I(\theta) = \lim_{n \rightarrow \infty} n^{-1} H(\theta)$ , sendo  $H(\theta)$  a matriz de informação observada, logo  $H(\theta)$  é um estimador consistente para  $I(\theta)$  (Lehmann e Casella, 1998).

A matriz de informação esperada da distribuição Beta parametrizada em termos de  $\mu$  e  $\phi$  é dada por:

$$I(\theta | \mathbf{y}) = \begin{bmatrix} I_{\mu\mu} & I_{\mu\phi} \\ I_{\phi\mu} & I_{\phi\phi} \end{bmatrix} \quad (8)$$

sendo:

$$\begin{aligned} I_{\mu\mu} &= n\phi^2\psi'(\mu\phi) + n\phi^2\psi'(\phi - \mu\phi), \\ I_{\phi\phi} &= n\mu^2\psi'(\mu\phi) + n(\mu - 1)^2\psi'(\phi - \mu\phi) - n\psi'(\phi), \\ I_{\mu\phi} &= n\mu\phi\psi'(\mu\phi) + n\psi(\mu\phi) - n(1 - \mu)\phi\psi'(\phi - \mu\phi) - n\psi(\phi - \mu\phi) - \sum_{i=1}^n \mathbb{E}[\log y_i] + \sum_{i=1}^n \mathbb{E}[\log(1 - y_i)] \end{aligned}$$

Após algumas manipulações é possível mostrar que

$$\sum_{i=1}^n \mathbb{E}[\log y_i] = n\psi(\mu\phi) - n\psi(\phi) \quad (9)$$

e

$$\sum_{i=1}^n \mathbb{E}[\log(1 - y_i)] = n \times \frac{1 + \phi(\mu - 1)[\pi \cot(\pi\phi(\mu - 1)) + \psi(\phi\mu - \phi) - \psi(\phi)]}{\phi(\mu - 1)} \quad (10)$$

em que  $\psi'(w) = d\phi(w)/dw$  é a função trigama e  $\cot(\cdot)$  representa a função cotangente.

Assim, os intervalos de confiança assintóticos para  $\mu$  e  $\phi$  são obtidos, respectivamente por

$$\hat{\mu} \pm z_{\alpha/2} \left[ \widehat{\mathbb{V}}(\hat{\mu}) \right]^{1/2} \quad \text{e} \quad \hat{\phi} \pm z_{\alpha/2} \left[ \widehat{\mathbb{V}}(\hat{\phi}) \right]^{1/2}$$

em que  $\mathbb{V}(\cdot)$  é o elemento da diagonal principal correspondente ao parâmetro da matriz  $I^{-1}(\hat{\theta})$  e  $z_{\alpha/2}$  é o percentil  $(1 - \alpha/2)$  da distribuição Normal padrão. É importante pontuar que a inversa de  $I(\theta)$  avaliada em  $\hat{\mu}$  e  $\hat{\phi}$  fornece a matriz assintótica de covariância dos estimadores de verossimilhança, isto é, fornece o erro padrão de  $\hat{\mu}$  e  $\hat{\phi}$ .

### 3 Testes de hipóteses

Nesta seção apresentamos a teoria dos testes que foram avaliados no estudo de simulação. Vale destacar que os quatro testes são fundamentados na teoria da verossimilhança, especificamente eles avaliam a função log-verossimilhança em diferentes escalas. A estatística da Razão de Verossimilhança compara a altura das log-verossimilhanças dos modelos completo e restrito, ao passo que a estatística de Wald compara a estimativa do parâmetro ( $\hat{\theta}$ ) com o valor do parâmetro sob a hipótese nula ( $\theta_0$ ). Enquanto que, a estatística Escore verifica a inclinação da log-verossimilhança sob a hipótese nula. Além disso, ambas as estatísticas Wald e Escore são formas quadráticas. Por outro lado, a estatística Gradiente não é uma forma quadrática e mede a projeção ortogonal do vetor escore avaliado em  $H_0$  sob o vetor  $(\hat{\theta} - \theta_0)$ .

Seja  $\mathbf{y} = (y_1, y_2, \dots, y_n)$  uma amostra aleatória de  $n$  observações independentes proveniente de (4), suponha o interesse em testar a hipótese nula  $H_0 : \theta = \theta_0$  versus a alternativa  $H_1 : \theta \neq \theta_0$ , em que  $\theta = (\mu, \phi)$  e  $\theta_0 = (\mu_0, \phi_0)$ .

Através do teste da razão de verossimilhança avaliamos o máximo das funções log-verossimilhanças dos modelos restrito e completo. A estatística do teste é dada por:

$$S_{LR} = 2 \left[ l(\hat{\theta} | \mathbf{y}) - l(\theta_0 | \mathbf{y}) \right] \quad (11)$$

Sob a hipótese nula,  $S_{LR}$  tem distribuição assintótica qui-quadrado com  $p = 2$  graus de liberdade (Pawitan, 2001).

Por outro lado, o teste de Wald consiste em comparar a estimativa do parâmetro ( $\hat{\theta}$ ) com o valor do parâmetro sob a hipótese nula ( $\theta_0$ ), por isso requer somente o ajuste do modelo completo. A estatística do teste tem forma quadrática definida por (Millar, 2011):

$$S_W = \left( \hat{\theta} - \theta_0 \right)^T I(\hat{\theta}) \left( \hat{\theta} - \theta_0 \right) \quad (12)$$

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

O teste Escore baseia-se na conjectura da função escore  $U(\theta_0)$  e da matriz de informação esperada  $I(\hat{\theta}_0)$  sob a hipótese nula. Conforme Millar (2011), a estatística do teste pode ser determinada em termos de forma quadrática por:

$$S_R = U(\theta_0)^T I^{-1}(\theta_0) U(\theta_0) \quad (13)$$

em que,  $U(\theta_0) = \left. \frac{\partial}{\partial \theta} l(\theta | y) \right|_{\theta=\theta_0}$  e  $I^{-1}(\theta_0)$  é a inversa da matriz de informação esperada sob a hipótese nula. A estatística  $S_R$ , sob a hipótese nula tem assintoticamente distribuição qui-quadrado com  $p = 2$  graus de liberdade. Nota-se que a estatística não requer a maximização da função log-verossimilhança em relação ao modelo completo, exigindo assim menor esforço computacional.

Proposta recentemente por Terrell (2002) a estatística gradiente assim como as estatísticas da Razão de Verossimilhança, Wald e Escore apresenta as mesmas propriedades assintóticas de primeira ordem. Além disso, sua forma é simples não envolvendo conhecimento sobre a matriz de informação. Dessa forma, dado o interesse em testar as hipóteses definidas a estatística gradiente é definida por

$$S_T = U(\theta_0)^T (\hat{\theta} - \theta_0) \quad (14)$$

Sendo que, sob a hipótese nula  $S_T$  tem distribuição assintótica qui-quadrado com  $p = 2$  graus de liberdade.

Nota-se, que assintoticamente estes quatros testes são equivalente, todavia como ressalta Buse (1982), Mazucheli e Louzada (2014) e Mazucheli et al. (2016) eles podem diferir em amostras finitas, principalmente em pequenas amostras. Vale ressaltar também que Muggeo e Lovison (2014) apresentam uma excelente interpretação geométrica e gráfica dessas estatísticas nas escalas das funções log-verossimilhança e escore.

## 4 Estudo de simulação e resultados

Diante dos testes descritos na seção anterior, que avaliam as hipóteses por meio de escalas distintas é de grande interesse verificar seu comportamento em distintas situações. Desse modo, foi conduzido um estudo de simulação Monte Carlo que averiguou o comportamento do tamanho do teste (*Erro do Tipo I*) em diferentes cenários. Assim, foram tomadas amostras de tamanhos  $n = 5, 10, 15, 20, 25, 30$  e  $50$  e parâmetros  $\mu = 0.05, 0.25, 0.5, 0.75$  e  $0.95$  e  $\phi = 5, 15$  e  $50$ . Para cada combinação  $(n, \mu, \phi)$  foram geradas  $M = 10000$  amostras pseudoaleatórias da distribuição Beta (4). Em todas as instâncias adotamos níveis de significância nominais  $\alpha = 0, 05$ .

O *Erro Tipo I* ( $\alpha$ ) foi estimado gerando amostras independentes, sob a hipótese nula, e calculando a proporção de vezes que  $H_0$  foi rejeitada erradamente, isto é,

$$\hat{\alpha} = \frac{\text{Número de vezes em que } H_0 \text{ é rejeitada} \mid H_0 \text{ é verdadeira}}{M} \quad (15)$$

Para atingir os objetivos propostos foi implementado no *software* SAS (SAS Institute Inc., 2011c) uma *macro* para calcular as probabilidades de rejeitar  $H_0$  quando a mesma foi verdadeira (valores- $p$ ) dos testes da Seção 3. Sendo que, o teste da Razão de Verossimilhança e Wald foram obtidos via PROC NLMIXED (SAS Institute Inc., 2011a). Já o teste Escore e Gradiente foram implementados no SAS/IML (SAS Institute Inc., 2011b). Abaixo apresenta-se o esquema utilizado na simulação. A macro do SAS, juntamente com instruções para seu uso, pode ser obtida solicitando para os autores.

```

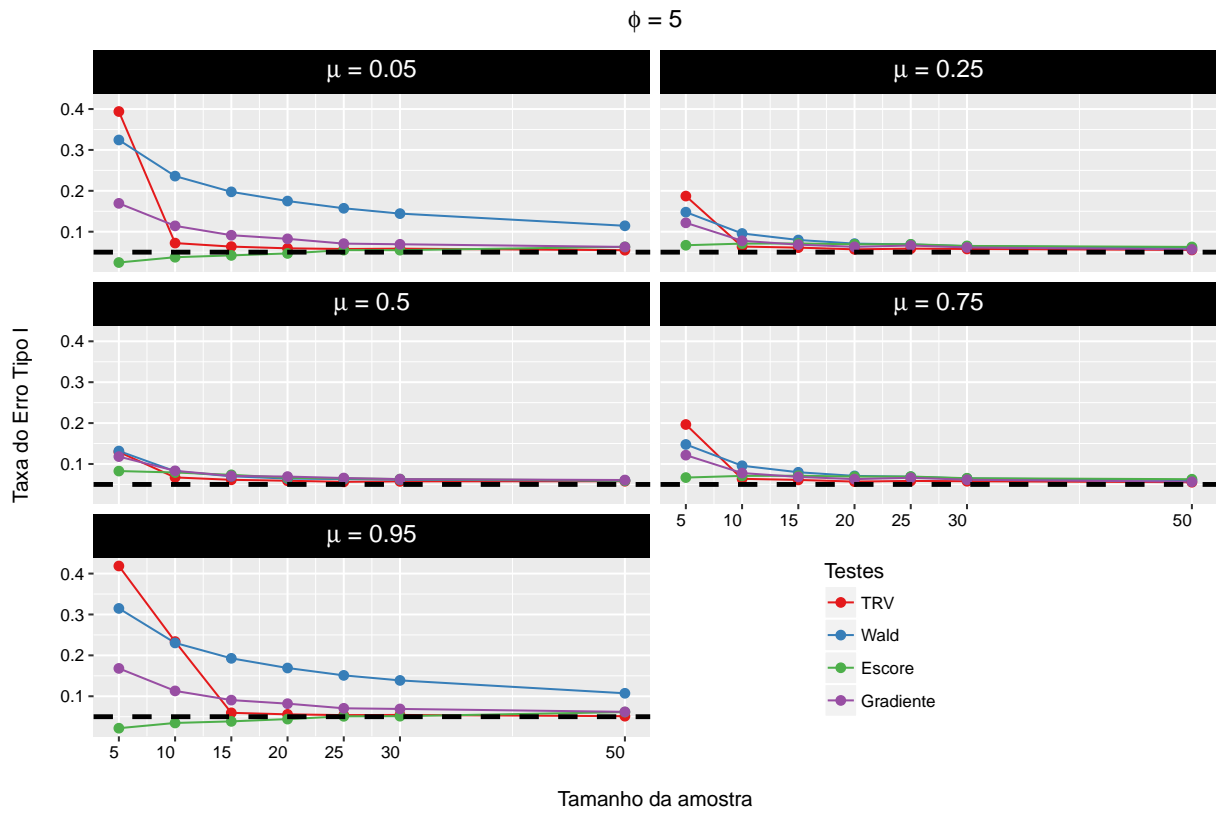
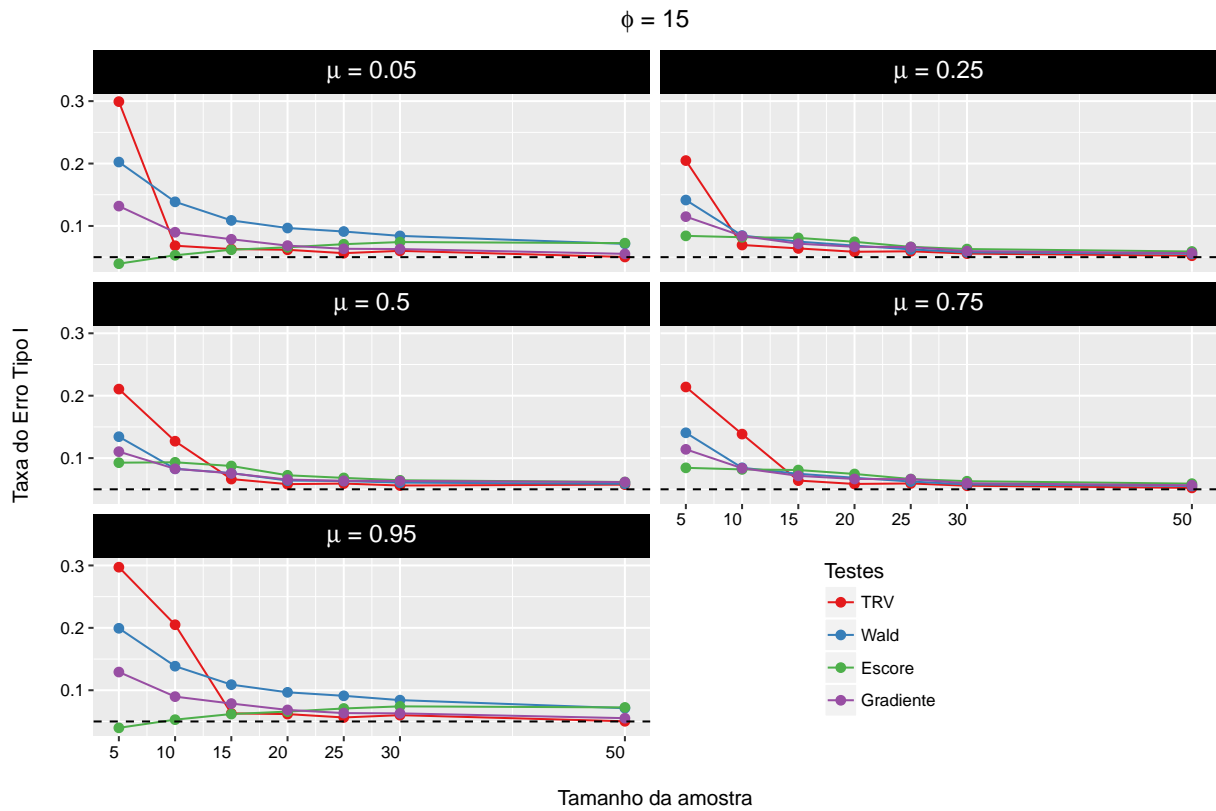
%macro simulacao_beta;
%do mu = 0.05 , 0.25, 0.5 , 0.75 , 0.95;
  %do phi = 5 , 15, 50;
    %do n=20 %to 100 %by 20;
      data simulacao;
      do b=1 to 10000;
        do i=1 to &n;
          y = valores aleatórios da Beta;
        end;end;
      run;
      proc nlmixed data=simulacao;
      -Teste 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;

```

Figura 2: Esquema

Os resultados obtidos com o processo de simulação podem ser observados nas figuras abaixo. No que diz respeito a figura 3, em que  $\phi = 5$  foi fixado, constatamos que para  $\mu = 0.05$  e  $0.95$  o teste de Wald excede o nível de significância nominal, mesmo quando temos uma amostra grande ( $n = 50$ ). Para os demais testes a probabilidade do *Erro Tipo I* convergiu para de  $0,05$  mais facilmente. Quando fixamos  $\phi = 15$  foi possível constatar que para os mesmo valores de  $\mu$  indicados anteriormente o teste de Wald, e neste caso, também o Escore, se mostram um pouco acima do nível de  $0,05$  para amostras grandes.

Por fim para  $\phi = 50$  notam-se uma considerável oscilação nas probabilidades do *Erro Tipo I* para o teste da Razão de Verossimilhança. Entretanto, quando a amostra atinge valores superiores a  $n = 30$  verifica-se uma estabilização da curva se aproximando do valor de interesse.

Figura 3: Resultados da simulação para  $\phi = 5$ Figura 4: Resultados da simulação para  $\phi = 15$

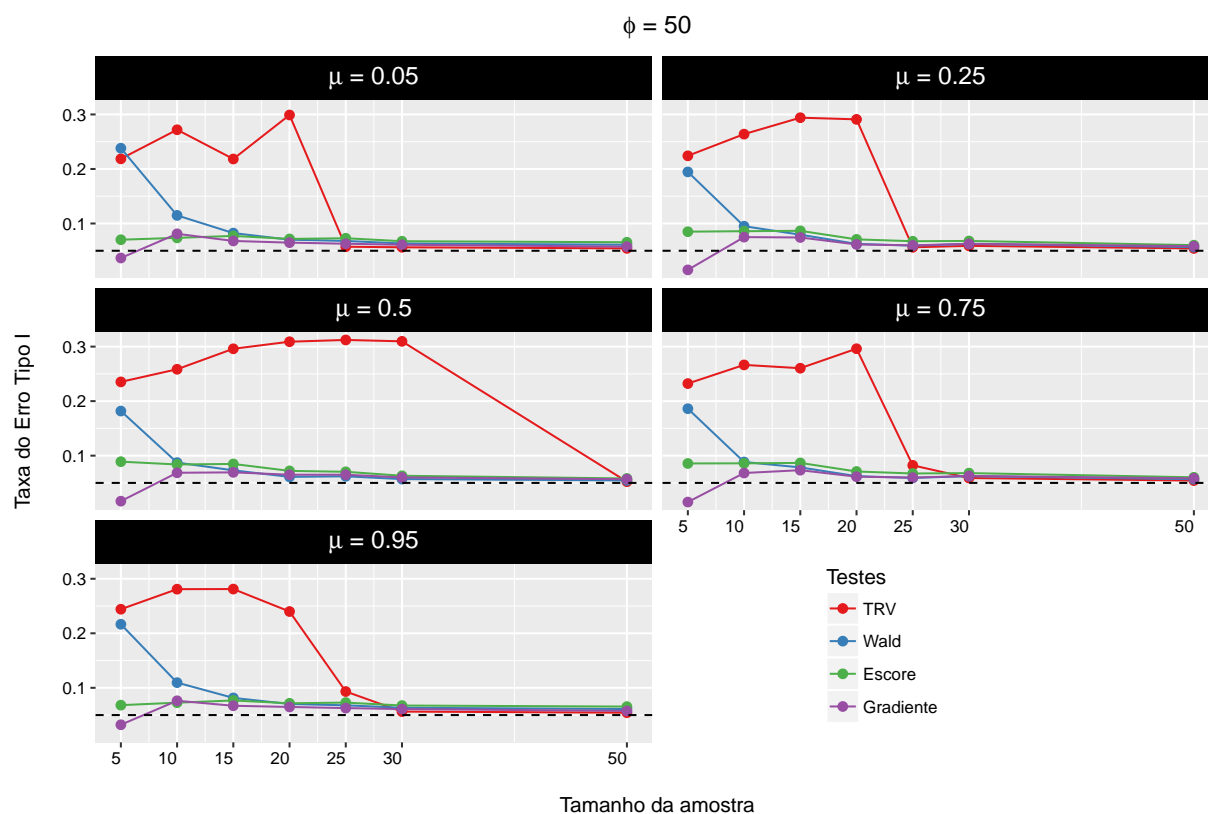


Figura 5: Resultados da simulação para  $\phi = 50$

## 5 Conclusões

Tendo em vista a flexibilidade da distribuição de probabilidade Beta para a modelagem importância de indicadores compreendidos em um intervalo. Este trabalho se ocupou em investigar as estatísticas dos testes da Razão de Verossimilhança, Wald, Escore e Gradiente, analisando a probabilidade do *Erro Tipo I* ( $\alpha$ ), em distintos cenários, nos quais, foram variados os parâmetros e o tamanho das amostras, com o objetivo de verificar o teste com o melhor desempenho. Dessa forma, no que tange os resultados, verificamos que todos os testes investigados se aproximam do nível de significância nominal de 0,05 conforme o tamanho da amostra aumenta. Além disso, constatou-se que o teste Gradiente, de um modo geral, foi o que apresentou o melhor desempenho quando comparado aos demais, principalmente para amostras menores que 20. Uma vez que na maior parte dos cenários analisados sua convergência para o valor de interesse se deu de forma mais acentuada. Por fim, chamamos a atenção para o cenário, no qual, o parâmetro de dispersão foi fixado em  $\phi = 50$ , neste caso, o teste da Razão de Verossimilhança se demonstrou instável para amostras pequenas, ou seja, para  $n < 30$ .....

## Referências

- Bhaumik, D., Kapur, K., Gibbons, R. (2009). Testing parameters of a Gamma distribution for small samples. *Technometrics*, 51(3), 326–334.
- Bury, K. (1999). *Statistical Distributions in Engineering*. Cambridge University Press.
- Buse, A. (1982). The Likelihood Ratio, Wald, and Lagrange Multiplier tests: An expository note. *The American Statistician*, 36(3a), 153–157.
- Caroni, C. (2010). Testing for the Marshall–Olkin extended form of the Weibull distribution. *Statistical Papers*, 51(2), 325–336.
- Ferrari, S. L. P., Cribari-Neto, F. (2004). Beta regression for modelling rates and proportions. *Journal of Applied Statistics*, 31(7), 799–815.
- Gupta, A. K., Nadarajah, S. (2004). *Handbook of Beta Distribution and Its Applications*. CRC Press.



- Johnson, N. L., Kotz, S., Balakrishnan, N. (1995). *Continuous Univariate Distributions*, vol 2. John Wiley & Sons Inc..
- Krit, M., Gaudoin, O., Xie, M., Remy, E. (2013). Simplified likelihood based goodness-of-fit tests for the Weibull distribution. *Communications in Statistics - Simulation and Computation*, 45(3), 920–951.
- Lehmann, E. J., Casella, G. (1998). *Theory of Point Estimation*. Springer Verlag.
- Mazucheli, J., Louzada, F. (2014). Discrimination between Odd Weibull and Weibull distributions. *Rev Bras Biom*, 32(2), 226–237.
- Mazucheli, J., Coelho-Barros, E. A., Louzada, F. (2016). On the hypothesis testing for the Weighted Lindley distribution. *Chilean Journal of Statistics*, 7(2).
- Millar, R. B. (2011). *Maximum likelihood estimation and inference: with examples in R, SAS and ADMB*, vol 1. John Wiley & Sons.
- Muggeo, V. M. R., Lovison, G. (2014). The “three plus one” likelihood-based test statistics: Unified geometrical and graphical interpretations. *The American Statistician*, 68(4), 302–306.
- Pawitan, Y. (2001). *In All Likelihood: Statistical Modelling and Inference Using Likelihood*. Oxford University Press, Oxford.
- SAS Institute Inc. (2011a). *The NLMIXED Procedure, SAS/STAT User’s Guide, Version 9.3*. Cary, NC: SAS Institute Inc., 4967–5062.
- SAS Institute Inc. (2011b). *SAS/IML 9.3 User’s Guide*. Cary, NC: SAS Institute Inc.
- SAS Institute Inc. (2011c). *SAS/STAT 9.3 User’s Guide*. Cary, NC: SAS Institute Inc.
- Terrell, G. R. (2002). The Gradient statistic. *Computing Science and Statistics*, 34(34), 206–215.