

# Relação Beta-Funcional entre o P-value e a Medida de Evidência Bayesiana na Comparação de duas Populações Normais

Luís Otávio Oliveira da Costa,

Edney Luís Oliveira Fernandes,

Maria Regina Madruga,

Héilton Ribeiro Tavares

Programa de Pós-Graduação em Matemática e Estatística, CCEN, UFPA,

Campus do Guamá, 66075-110, Belém, PA

E-mail: looc@ufpa.br, edneyluis@yahoo.com.br,

madruga@ufpa.br, heliton@ufpa.br

## 1 Introdução

O problema estatístico de *Teste de Hipóteses* consiste em rejeitar ou não uma hipótese  $H_0$  sobre o valor de um parâmetro desconhecido com base na informação trazida pela amostra. Na abordagem Clássica é bastante usada uma medida de evidência denominada *p-value*, cujo cálculo baseia-se na distribuição amostral da estatística do teste. Na abordagem Bayesiana, as medidas de evidência usuais para teste de hipóteses são o *Fator de Bayes* e a *Probabilidade Posterior de  $H_0$* . Alguns autores, Berger e Selke (1987) e Berger e Delampady (1987), entre outros, apresentaram e discutiram os conflitos entre o p-value e as medidas de evidência Bayesianas, alertando para o fato de que em algumas situações o *p-value* pode não ser uma boa medida de evidência para uma hipótese estatística precisa (hipóteses com medida de Lebesgue nula). Com isso, houve o surgimento de novas medidas de evidência para teste de hipóteses precisas. Neste trabalho será apresentada a medida de evidência Bayesiana proposta por Pereira e Stern (1999), cujo procedimento é denominado **Full Bayesian Significance Test (FBST)**. O objetivo é comparar, através de resultados de simulações, o *p-value* da abordagem Clássica, usando o *Teste da Razão de Verossimilhança Generalizada (TRVG)* (Bolfarine e Sandoval (2001)) e a medida proposta por Pereira e Stern (1999), sugerindo uma possível relação funcional entre essas duas medidas para testar a igualdade das médias de duas populações normais. Resultados empíricos preliminares indicam que a medida Bayesiana pode ser obtida pela função de distribuição acumulada do modelo Beta, calculada no p-value correspondente.

Pereira e Stern (1999) introduziram a seguinte medida de evidência em favor de uma hipótese precisa,

**Definição 1.** *Considere um modelo estatístico paramétrico, i.e., uma quintupla  $(\mathcal{X}, \mathcal{A}, \mathbf{F}, \Theta, \pi)$ , onde  $\mathcal{X}$  é um espaço amostral,  $\mathcal{A}$  é uma sigma-álgebra conveniente de subconjuntos de  $\mathcal{X}$ ,  $\mathbf{F}$  é uma classe*

*de distribuições de probabilidade em  $\mathcal{A}$  indexadas no espaço paramétrico  $\Theta$  e  $\pi$  é uma densidade a priori em (uma sigma-álgebra de)  $\Theta$ . Suponha que um subconjunto  $\Theta_0$  de  $\Theta$  tendo medida de Lebesgue nula (com respeito a  $\Theta$ ) é de interesse. Seja  $\pi(\theta|\mathbf{x})$  uma densidade posterior de  $\theta$ , dada a observação amostral  $\mathbf{x}$ , e  $T(\mathbf{x}) = \{\theta \in \Theta : \pi(\theta|\mathbf{x}) > \sup_{\Theta_0} \pi(\theta|\mathbf{x})\}$ . A **medida de evidência de Pereira-Stern** é definida como  $EV(\Theta_0, \mathbf{x}) = 1 - Pr[\theta \in T(\mathbf{x})|\mathbf{x}]$  e um teste (ou procedimento) de Pereira-Stern é aceitar  $\Theta_0$  sempre que  $EV(\Theta_0, \mathbf{x})$  é “grande”.*

Como podemos ver da Definição 1, a medida de evidência de Pereira-Stern considera, em favor de uma hipótese precisa, todos os pontos do espaço paramétrico cujos valores da densidade posterior são, no máximo, tão grandes quanto seu supremo em  $\Theta_0$ ; falando grosseiramente, considera todos os pontos que são menos “prováveis” do que algum ponto em  $\Theta_0$ . De acordo com Pereira e Stern (1999), um valor grande da  $EV(\Theta_0, \mathbf{x})$  significa que o subconjunto  $\Theta_0$  cai em uma região do espaço paramétrico de alta probabilidade posterior, portanto, os dados suportam a hipótese nula; por outro lado, um valor pequeno da  $EV(\Theta_0, \mathbf{x})$  indica que  $\Theta_0$  está em uma região do espaço paramétrico de baixa probabilidade posterior, portanto, os dados nos levariam a desacreditar da hipótese nula. Uma vantagem deste procedimento é que ele contorna a dificuldade de tratar com uma hipótese precisa, pois não há necessidade de introduzir uma probabilidade positiva a priori como no teste Bayesiano padrão (Jeffreys (1961)). Pereira e Stern (1999) defendem que o uso de  $EV(\Theta_0, \mathbf{x})$  para avaliar a evidência trazida pelos dados para  $\Theta_0$  é um procedimento “Bayesiano”, uma vez que apenas a densidade posterior está envolvida. Madruga et al. (2001) apresentam funções de perda que tornam o procedimento de Pereira-Stern um legítimo procedimento “Bayesiano”, pois devem ser chamados “Bayesianos” apenas os procedimentos que minimizam funções perdas esperadas, a solução coerente para o problema de decisão. As-

sim, o teste de Pereira-Stern consiste em

- **Rejeitar**  $H_0$  se  $EV(\Theta_0, \mathbf{x}) \leq K$
- **Aceitar**  $H_0$  se  $EV(\Theta_0, \mathbf{x}) > K$

onde  $K$  é um ponto crítico cujo valor depende da função de perda escolhida.

Por outro lado, o TRVG também pode ser usado para testar uma hipótese precisa, e a regra de decisão é baseada na distribuição assintótica da estatística

$$-2 \log \lambda(\mathbf{x}) = -2 \log \left( \frac{\sup_{\theta \in \Theta_0} L(\theta; \mathbf{x})}{\sup_{\theta \in \Theta} L(\theta; \mathbf{x})} \right),$$

onde  $L(\theta; \mathbf{x})$  é a *função de verossimilhança* dos dados amostrais. A distribuição assintótica é qui-quadrado com  $\nu$  graus de liberdade ( $\chi_\nu^2$ ), e  $\nu$  é dado pela dimensão de  $\Theta \setminus \Theta_0$ . O cálculo do *p-value* ( $p$ ) é dado pela seguinte probabilidade,  $p = Pr\{\chi_\nu^2 > -2 \log \lambda(\mathbf{x})\}$ .

Resultados empíricos preliminares em alguns testes conhecidos (Madruga (2002)) indicam que a  $EV(\Theta_0, \mathbf{x})$  pode ser obtida como a função de distribuição acumulada de um modelo de probabilidades Beta com parâmetros  $a$  e  $b$ , calculada no *p-value* correspondente, ou seja,

$$EV(\Theta_0, \mathbf{x}) = \int_0^p \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} u^{a-1}(1-u)^{b-1} du. \quad (1)$$

As estimativas de  $a$  e  $b$  no ajuste da relação (1) variam nos testes já estudados. Nosso objetivo é estudar essa variação como função dos tamanhos amostrais e da variância comum conhecida para o problema de comparar as médias de duas populações Normais.

## 2 Comparação das médias de duas Normais

Com base em amostras obtidas das duas populações,  $X_1, \dots, X_n$  uma amostra aleatória de  $X \sim N(\mu_1, \sigma^2)$ ,  $Y_1, \dots, Y_m$  uma amostra aleatória de  $Y \sim N(\mu_2, \sigma^2)$ , e admitindo  $\sigma^2$  conhecida, queremos testar as hipóteses  $H_0 : \mu_1 = \mu_2$  contra  $H_1 : \mu_1 \neq \mu_2$ , ou seja, tem-se

$$\Theta = \{(\mu_1, \mu_2) : \mu_1 \in \mathbb{R} \text{ e } \mu_2 \in \mathbb{R}\}$$

$$\Theta_0 = \{(\mu_1, \mu_2) \in \Theta : \mu_1 = \mu_2\}.$$

A função de verossimilhança dos dados observados  $(\mathbf{x}, \mathbf{y})$ , com  $\mathbf{x} = \{x_1, \dots, x_n\}$  e  $\mathbf{y} = \{y_1, \dots, y_m\}$ , é dada por

$$L(\theta; \mathbf{x}, \mathbf{y}) = \left( \frac{1}{2\pi\sigma^2} \right)^{\frac{n+m}{2}} \exp \left\{ -\frac{1}{2\sigma^2} [\delta_1 + \delta_2] \right\}, \quad (2)$$

com  $\delta_1 = (n-1)S_x^2 + n(\mu_1 - \bar{x})^2$  e  $\delta_2 = (m-1)S_y^2 + m(\mu_2 - \bar{y})^2$ , sendo  $\bar{x} = (1/n) \sum_{i=1}^n x_i$  e  $\bar{y} = (1/m) \sum_{j=1}^m y_j$  as médias amostrais,  $S_x^2 = (1/(n-1)) \sum_{i=1}^n (x_i - \bar{x})^2$  e  $S_y^2 = (1/(m-1)) \sum_{j=1}^m (y_j - \bar{y})^2$  as variâncias amostrais.

### • Construção da $EV(\Theta_0; \mathbf{x}, \mathbf{y})$ usando o procedimento FBST:

Considerando a priori imprópria usual para  $\theta = (\mu_1, \mu_2)$ ,  $\pi(\theta) = c$ , temos que a densidade posterior é dada por

$$\pi(\theta|\mathbf{x}, \mathbf{y}) \propto \exp \left\{ -\frac{1}{2\sigma^2} [n(\mu_1 - \bar{x})^2 + m(\mu_2 - \bar{y})^2] \right\}, \quad (3)$$

que pode ser fatorada como o produto das densidades posteriores de  $\mu_1$  e  $\mu_2$ ,  $\pi(\theta|\mathbf{x}, \mathbf{y}) = \pi_1(\mu_1|\mathbf{x}) \times \pi_2(\mu_2|\mathbf{y})$ , com as funções  $\pi_i(\cdot)$ ,  $i = 1, 2$ , representando, respectivamente, as f.d.p.'s dos seguintes modelos de probabilidade:  $\mu_1|\mathbf{x} \sim N(\bar{x}; \sigma^2/n)$  e  $\mu_2|\mathbf{y} \sim N(\bar{y}; \sigma^2/m)$ . Para o cálculo da  $EV(\Theta_0; \mathbf{x}, \mathbf{y})$ , primeiramente maximizamos a distribuição posterior  $\pi(\theta|\mathbf{x}, \mathbf{y})$  sob a hipótese nula, obtendo

$$\pi(\hat{\mu}|\mathbf{x}, \mathbf{y}) \propto \exp \left\{ -\frac{1}{2\sigma^2} [n(\hat{\mu} - \bar{x})^2 + m(\hat{\mu} - \bar{y})^2] \right\},$$

com  $\hat{\mu} = (n\bar{x} + m\bar{y})/(n+m)$ . Em seguida, resolvemos a seguinte integral

$$I = \int_{T(\mathbf{x}, \mathbf{y})} \pi(\theta|\mathbf{x}, \mathbf{y}) d\theta$$

com  $T(\mathbf{x}, \mathbf{y}) = \{\theta \in \Theta : \pi(\theta|\mathbf{x}, \mathbf{y}) > \pi(\hat{\mu}|\mathbf{x}, \mathbf{y})\}$ . A integral acima pode ser expressa como a esperança posterior da função indicadora  $h(\theta) = 1(\theta \in T(\mathbf{x}, \mathbf{y}))$ , ou seja,

$$I = E[h(\theta)|\mathbf{x}, \mathbf{y}] = \int_{\Theta} 1(\theta \in T(\mathbf{x}, \mathbf{y})) \pi(\theta|\mathbf{x}, \mathbf{y}) d\theta, \quad (4)$$

que pode ser aproximada pelo Método de Monte Carlo. Para isso, simulamos uma amostra aleatória de tamanho  $M$ ,  $(\theta_1, \dots, \theta_M)$ , da distribuição em (3) e aproximamos a integral (4) por

$$\hat{I} = \frac{1}{M} \sum_{j=1}^M h(\theta_j).$$

Logo, a medida de evidência do procedimento FBST é aproximada por

$$EV(\Theta_0; \mathbf{x}, \mathbf{y}) \approx 1 - \frac{1}{M} \sum_{j=1}^M h(\theta_j). \quad (5)$$

### Construção do p-value usando o TRVG:

A estatística do TRVG é dada por

$$-2 \log \lambda(\mathbf{x}, \mathbf{y}) = \frac{1}{\sigma^2} [n(\hat{\mu} - \bar{x})^2 + m(\hat{\mu} - \bar{y})^2],$$

e usando sua distribuição assintótica  $\chi_1^2$ , obtemos o p-value

$$p = \int_{-2 \log \lambda(\mathbf{x}, \mathbf{y})}^{+\infty} \frac{(1/2)^{1/2}}{\Gamma(1/2)} u^{\frac{1}{2}-1} e^{-\frac{u}{2}} du.$$

## 3 Resultados

Foi utilizado o aplicativo MATLAB (The Mathworks, Inc. [1999]) para a obtenção dos resultados. A Tabela 1 apresenta o p-value ( $p$ ) e a  $EV(\Theta_0; \mathbf{x}, \mathbf{y})$  para 20 iterações, em que as amostras foram geradas de duas distribuições normais, com  $n = m = 30$ ,  $\sigma^2 = 9$ ,  $\mu_1 = 0$  e diferentes valores de  $\mu_2$ . A Figura 1 apresenta o gráfico de dispersão dos pares  $(p, EV(\Theta_0; \mathbf{x}, \mathbf{y}))$  apresentados na Tabela 1.

$ \mu_1 - \mu_2 $	$EV(\Theta_0; \mathbf{x}, \mathbf{y})$	$p$
0,0	0,9642	0,7835
0,1	0,7653	0,4608
0,2	0,9820	0,8585
0,3	0,5939	0,3052
0,4	0,7234	0,4164
0,5	0,7064	0,4055
0,6	0,9966	0,9294
0,7	0,9465	0,7348
0,8	0,1328	0,0434
0,9	0,0437	0,0124
1,0	0,9857	0,8696
1,2	0,0010	0,0002
1,4	0,1896	0,0683
1,6	0,2517	0,0981
2,0	0,0077	0,0018
2,5	0,0038	0,0010
3,0	0,0000	0,0000
3,5	0,0001	0,0000
4,0	0,0000	0,0000
4,5	0,0000	0,0000

Tabela 1: Medidas de Evidência para o teste de igualdade das médias de duas populações normais

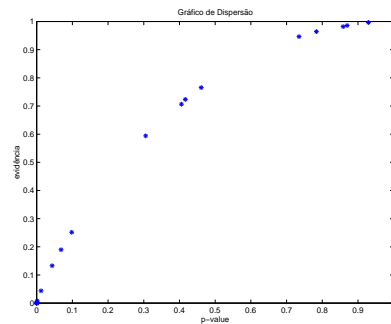


Figura 1: Gráfico de dispersão entre as medidas de evidência

Ajustando o modelo (1) aos pontos da Figura 1, obtivemos as estimativas  $\hat{a} = 0,8423$  e  $\hat{b} = 2,0468$ .

Para estudarmos a variação nas estimativas de  $a$  e  $b$  no ajuste da relação (1), como função dos tamanhos amostrais  $n$  e  $m$  e da variância  $\sigma^2$ , foram feitas algumas simulações. A Tabela 2 apresenta as estimativas de  $a$  e  $b$  para  $\sigma^2 = 9$  e diferentes valores comuns para  $n$  e  $m$ . A Tabela 3 apresenta as estimativas de  $a$  e  $b$  para  $n = m = 30$  e diferentes valores para  $\sigma^2$ . Os resultados apresentados nas Tabelas 2 e 3 sugerem que as estimativas de  $a$  e  $b$  no ajuste da relação (1) não dependem das dimensões amostrais e nem da variância.

Valores para $n$ e $m$	$a$	$b$
05	0,84	2,04
10	0,85	2,05
15	0,83	2,03
20	0,84	2,05
25	0,85	2,07
30	0,84	2,05
35	0,83	2,01
40	0,83	2,03
45	0,83	2,03
50	0,85	2,04
55	0,83	2,00
60	0,84	2,03
65	0,84	2,04
70	0,84	2,04
100	0,83	2,01
200	0,85	2,07
300	0,85	2,07
600	0,84	2,02

Tabela 2: Estimativas de  $a$  e  $b$  para diferentes valores de  $n$  e  $m$  e  $\sigma^2 = 9$

Valores para $\sigma^2$	$a$	$b$
01	0,83	2,01
03	0,83	2,03
05	0,85	2,07
07	0,84	2,05
09	0,84	2,05
11	0,85	2,06
13	0,84	2,06
15	0,84	2,05
17	0,85	2,05
19	0,84	2,04
21	0,84	2,05
23	0,85	2,05
25	0,85	2,05
27	0,85	2,06
29	0,85	2,05
70	0,84	2,04
90	0,84	2,06
200	0,84	2,05

Tabela 3: Estimativas de  $a$  e  $b$  para  $n = m = 30$  e diferentes valores de  $\sigma^2$

## Referências

- [1] J. O. Berger e T. Selke, Testing a Point Null Hypothesis: The Irreconcilability of P values and Evidence, *Journal of the American Statistical Association*, 82 (1987) 112-139.
- [2] J. O. Berger e M. Delampady, Testing Precise Hypotheses, *Statistical Science*, 2 (1987) 317-352.
- [3] H. Bolfarine e M. C. Sandoval, "Introdução à Inferência Estatística", Sociedade Brasileira de Matemática, Rio de Janeiro, 2001.

- [4] H. Jeffreys, "Theory of Probability", University Press, Oxford, 1961.
- [5] M. R. Madruga, L. G. Esteves e S. Wechsler, On the Bayesianity of Pereira-Stern tests, *Test*, 10 (2001) 291-299.
- [6] M. R. Madruga, Teste de Significância: uma proposta genuinamente Bayesiana. 2002. 77 f. Tese (Doutorado em Estatística) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo.
- [7] The Mathworks, Inc. (1999). *MATLAB - The Language of Technical Computing*. Version 5.3.0.10183(R11).
- [8] C. A. de B. Pereira and J. Stern, Evidence and Credibility: a full Bayesian test of precise hypothesis, *Entropy*, 1 (1999) 99-110.