TESTES DE IGUALDADE E DE COMPARAÇÕES MÚLTIPLAS PARA VÁRIAS PROPORÇÕES BINOMIAIS INDEPENDENTES

Nádia Giaretta BIASE[1]
Daniel Furtado FERREIRA[2]

▪ RESUMO: Na teoria frequentista, inferências sobre mais que duas proporções binomiais são realizadas utilizando-se a análise de variância e os procedimentos de comparações múltiplas ou, ainda, por meio dos testes assintóticos e de métodos de computação intensiva. Todos esses procedimentos apresentam uma limitação, que pode ser atribuída à violação das pressuposições exigidas pelos testes. Por essa razão, este trabalho foi realizado com o objetivo de propor uma abordagem bayesiana para realizar um teste global de igualdade de várias proporções binomiais e um teste de comparações múltiplas de proporções binomiais. O desempenho desses testes foi avaliado utilizando simulação Monte Carlo. Foram geradas k populações binomiais independentes com parâmetros π_i e n_i, $i = 1, 2, \cdots, k$ e simuladas amostras de Monte Carlo para cada configuração envolvendo combinações das quantidades k, n_i's e π_i's, considerando *prioris* conjugadas betas com parâmetros α_i e β_i, fixados por tentativa e erro, buscando minimizar as taxas de erro tipo I e maximizar o poder. O TB proposto para a igualdade de várias proporções binomiais apresentou excelente desempenho no controle do erro tipo I e poder relativamente alto. O TCMB para proporções binomiais, sob H_0 completa e parcial, foi conservativo e apresentou grande poder.

▪ PALAVRAS-CHAVE: Simulação Monte Carlo; teste bayesiano de igualdade de várias proporções binomiais; teste de comparações múltiplas bayesianas; erro tipo I; poder.

1 Introdução

A comparação de várias proporções binomiais é relevante em muitos estudos científicos. Na inferência clássica, essa comparação é normalmente realizada

[1]Universidade Federal de Uberlândia – UFU, Faculdade de Matemática – FAMAT, Campus Santa Mônica, CEP: 38408-100, Uberlândia, MG, Brasil. E-mail: *nadia@famat.ufu.br*
[2]Universidade Federal de Lavras – UFLA, Departamento de Ciências Exatas – DEX, Caixa Postal 37, CEP: 37200-000, Lavras, MG, Brasil. E-mail: *danielff@dex.ufla.br*

aplicando-se, inicialmente, um teste F para a igualdade de todas as proporções binomiais em uma análise de variância, seguida da aplicação de um teste de comparações múltiplas para averiguar quais proporções binomiais diferem entre si, caso a hipótese de nulidade global seja rejeitada.

Um problema encontrado pelos pesquisadores com a aplicação dessa técnica é que, no caso das proporções binomiais, as pressuposições de normalidade dos resíduos e a homogeneidade de variâncias não são atendidas. Nesse caso, métodos alternativos podem ser utilizados para amenizar o problema. Dentre esses métodos, Nelder e Wedderburn (1972) introduziram uma modelagem estatística de dados, utilizando os modelos lineares generalizados, que envolve uma variedade de modelos pertencentes à família exponencial, incluindo o modelo binomial. No entanto, uma limitação desse método está relacionada com as distribuições das estatísticas dos testes, que são apenas assintóticas (McCulloch e Searle, 2001; Dobson e Barnett, 2008). Procedimentos de comparações múltiplas baseados na família exponencial não são encontrados.

Para testar a hipótese de nulidade global de igualdade de várias proporções binomiais, outros métodos alternativos baseiam-se na teoria assintótica. Williams (1988), Krishnamoorthy e Peng (2008) e Biase e Ferreira (2009) avaliaram o desempenho de alguns desses testes, dos quais podem-se citar o teste de razão de verossimilhanças e o de X^2 de Pearson. Especificando um valor de referência π_0, Krishnamoorthy *et al.* (2004) também desenvolveram um estudo para testar várias proporções binomiais e, nos casos em que a hipótese de nulidade foi rejeitada, apresentaram um procedimento de construção de intervalos de confiança simultâneos para proporções binomiais π_i, $i = 1, 2, \cdots, k$ para identificar quais dessas proporções causaram a rejeição da hipótese nula, sendo k o número de populações binomiais independentes.

Em relação as comparações múltiplas de proporções binomiais, Piegorsch (1991) considerou intervalos de confiança simultâneos, envolvendo qualquer conjunto finito de contrastes, usando a aproximação de Bonferroni aplicada ao intervalo de confiança de Wald e, intervalos de confiança simultâneos, incluindo todos os pares de comparações ($\pi_i - \pi_{i'}$, $i \neq i' = 1, 2, \cdots, k$), utilizando o intervalo de Wald juntamente com a distribuição da amplitude normal padronizada. As taxas de erro desses dois procedimentos foram consideravelmente maiores do que o valor nominal para pequenas amostras e, portanto, apresentaram desempenhos ruins. Esse autor mostrou também que um melhor desempenho desses intervalos foi obtido empregando a formulação de Jeffreys-Perks, motivada pela aproximação bayesiana de Beal (1987). Quando implementado para as comparações múltiplas usando a distribuição padronizada, com tamanhos amostrais iguais, $n_i = n$, o intervalo para a diferença entre as proporções, $\pi_i - \pi_{i'}$, é dado por: $(1 + d^2)^{-1}[(\hat{\pi}_i - \hat{\pi}_{i'}) \pm d\{(2 - \tilde{\theta}_{ij})\tilde{\theta}_{ij}(1 + d^2 - (\hat{\pi}_i - \hat{\pi}_{i'})^2)\}^{1/2}]$, em que $\tilde{\theta}_{ij} = \{n(\hat{\pi}_i - \hat{\pi}_{i'}) + 1\}/(n+1)$ e $d = Q_{k,\infty}(\alpha)/2\sqrt{n}$, sendo que $Q_{k,\infty}(\alpha)$ representa o quantil superior $100\alpha\%$ da distribuição normal padrão. Com essas modificações, obteve-se probabilidades de cobertura próximas ao valor nominal, para amostras pequenas e moderadas.

Biase e Ferreira (2006) propuseram testes de comparações múltiplas para parâmetros binomiais utilizando métodos de reamostragem *bootstrap* e, obtiveram resultados de alta qualidade. Biase e Ferreira (2009) também desenvolveram um teste de comparações múltiplas assintóticas de proporções binomiais, baseado em formas quadráticas, que apresentou excelentes resultados.

Por outro lado, testes de hipóteses globais de igualdade de várias proporções binomiais e testes de comparações múltiplas de proporções binomiais podem ser realizados utilizando-se uma abordagem bayesiana. Essa abordagem permite incorporar o conhecimento *a priori* dos pesquisadores sobre os parâmetros de interesse. Não foram encontrados trabalhos que aplicassem qualquer um desses testes para proporções binomiais. Entretanto, existem trabalhos que utilizam procedimentos bayesianos para realizar comparações múltiplas em populações normais (Shaffer, 1999; Bratcher e Hamilton, 2005; Andrade, 2008). Esse último autor realizou comparações múltiplas em populações normais homocedásticas e heterocedásticas. A partir da distribuição *a posteriori*, Andrade (2008) considerou, sob H_0, k cadeias de médias, assumindo médias constantes, e obteve a amplitude padronizada da *posteriori*, por meio da distribuição *a posteriori* das médias. Para realizar as inferências, sob H_0, obteve-se a diferença mínima significativa e o intervalo de credibilidade, que apresentou excelentes desempenhos. No contexto de tabelas de contingência, Agresti e Min (2005) também avaliaram o desempenho de intervalos de credibilidade bayesiano para proporções binomiais via simulação Monte Carlo. Ressaltaram a importância de se usar tais métodos e os recomendaram para validação de testes bayesianos.

Assim, este trabalho foi realizado com o objetivo de propor um teste bayesiano para a hipótese de nulidade global de igualdade de várias proporções binomiais e um teste de comparações múltiplas bayesiano, avaliando o desempenho desses testes por meio de simulação Monte Carlo.

2 Metodologia

Para testar a hipótese de igualdade de k proporções binomiais independentes, foi proposto, neste trabalho, um teste bayesiano usando *prioris* betas conjugadas. Numa segunda etapa, desenvolveu-se um teste bayesiano para realizar comparações múltiplas. Em ambos os casos, o desempenho foi avaliado por simulação Monte Carlo, conforme justificativa apresentada por Agresti e Min (2005).

2.1 Teste bayesiano para igualdade de proporções binomiais

Para realizar o teste da hipótese de igualdade de k proporções binomiais independentes $H_0 : \pi_1 = \pi_2 = \cdots = \pi_k = \pi_0$, foi proposto o método bayesiano descrito na sequência. Inicialmente, são consideradas k amostras aleatórias independentes, de tamanhos $n_1, n_2, \cdots, n_k$, de populações binomiais com probabilidade de sucesso dos eventos de interesse, dados por $\pi_1 = \pi_2 = \cdots = \pi_k$, respectivamente. Sejam $y_1, y_2, \cdots, y_k$ as realizações do número de sucesso nas amostras das populações 1, 2, $\cdots$, k, respectivamente, então, a função de verossimilhança é dada por:

$$L(\boldsymbol{y}|\boldsymbol{\pi}) = \prod_{i=1}^{k} \frac{n_i!}{y_i!(n_i - y_i)!} \pi_i^{y_i} (1 - \pi_i)^{n_i - y_i}, \tag{1}$$

em que: $\boldsymbol{y} = [\, y_1,\ y_2,\ \ldots,\ y_k \,]^\top$ é o vetor das realizações do número de sucesso em cada uma das k populações e $\boldsymbol{\pi} = [\, \pi_1,\ \pi_2,\ \ldots,\ \pi_k \,]^\top$ é o vetor de parâmetros das k populações binomiais independentes.

Optou-se por escolher *prioris* conjugadas. Logo, a distribuição de cada componente de $\boldsymbol{\pi}$ foi uma beta com parâmetros α_i e β_i. Como, os π's são independentes, a distribuição conjunta *a priori* de $\boldsymbol{\pi}$ foi:

$$p(\boldsymbol{\pi}) \propto \prod_{i=1}^{k} \pi_i^{\alpha_i - 1} (1 - \pi_i)^{\beta_i - 1}. \tag{2}$$

A distribuição conjunta *a posteriori* foi obtida multiplicando-se (1) por (2) e o resultado é:

$$\begin{aligned} p(\boldsymbol{\pi}|\boldsymbol{y}) &\propto p(\boldsymbol{\pi}) L(\boldsymbol{y}|\boldsymbol{\pi}) \\ &\propto \prod_{i=1}^{k} \pi_i^{\alpha_i - 1} (1 - \pi_i)^{\beta_i - 1} \prod_{i=1}^{k} \frac{n_i!}{y_i!(n_i - y_i)!} \pi_i^{y_i} (1 - \pi_i)^{n_i - y_i} \\ &\propto \pi_1^{\alpha_1 + y_1 - 1} (1 - \pi_1)^{\beta_1 + n_1 - y_1 - 1} \ldots \pi_k^{\alpha_k + y_k - 1} (1 - \pi_k)^{\beta_k + n_k - y_k - 1}, \end{aligned}$$

logo, a distribuição conjunta *a posteriori* de $\boldsymbol{\pi}$ é:

$$\boldsymbol{\pi}|\boldsymbol{y} \sim \prod_{i=1}^{k} B_i(\alpha_i + y_i, \beta_i + n_i - y_i),$$

em que $B_i(\alpha, \beta)$ refere-se à distribuição beta com parâmetros α e β. Assim, para cada π_i, $i = 1,\, 2,\, \cdots,\, k$ a distribuição é:

$$\pi_i|\boldsymbol{y} \sim B_i(\alpha_i + y_i, \beta_i + n_i - y_i).$$

Assim, a média e a variância da distribuição *a posteriori* de π_i são:

$$E(\pi_i|\boldsymbol{y}) = \frac{\alpha_i + y_i}{\alpha_i + \beta_i + n_i} = \mu_{\pi_i}, \tag{3}$$

$$V(\pi_i|\boldsymbol{y}) = \frac{(\alpha_i + y_i)(\beta_i + n_i - y_i)}{(\alpha_i + \beta_i + n_i)^2 (\alpha_i + \beta_i + n_i + 1)} = \sigma^2_{\pi_i}. \tag{4}$$

Sob a hipótese nula, $H_0 : \pi_1 = \pi_2 = \cdots = \pi_k = \pi$, a função de verossimilhança foi simplificada da seguinte forma:

$$\begin{aligned} L(\boldsymbol{y}|\pi) &\propto \pi^{y_1} (1 - \pi)^{n_1 - y_1} \pi^{y_2} (1 - \pi)^{n_2 - y_2} \cdots \pi^{y_k} (1 - \pi)^{n_k - y_k} \\ &\propto \pi^{\sum_{i=1}^{k} y_i} (1 - \pi)^{\sum_{i=1}^{k} n_i - \sum_{i=1}^{k} y_i}. \end{aligned} \tag{5}$$

A distribuição *a priori* para π, comum a todas as populações, foi escolhida como uma beta com hiperparâmetros α e β. Assim,

$$P(\pi) \propto \pi^{\alpha-1}(1-\pi)^{\beta-1}. \tag{6}$$

A distribuição *a posteriori* de π foi obtida multiplicando-se (5) por (6), da seguinte forma:

$$\begin{aligned} P(\pi|\boldsymbol{y}) &\propto P(\pi)L(\boldsymbol{y}|\pi) \\ &\propto \pi^{\alpha-1}(1-\pi)^{\beta-1}\pi^{\sum_{i=1}^{k} y_i}(1-\pi)^{\sum_{i=1}^{k} n_i - \sum_{i=1}^{k} y_i} \\ &\propto \pi^{\alpha+\sum_{i=1}^{k} y_i - 1}(1-\pi)^{\beta+\sum_{i=1}^{k} n_i - \sum_{i=1}^{k} y_i - 1}. \end{aligned}$$

Logo, a distribuição *a posteriori* de $\pi|\boldsymbol{y}$ foi:

$$\pi|\boldsymbol{y} \sim B\left(\alpha + \sum_{i=1}^{k} y_i, \beta + \sum_{i=1}^{k} n_i - \sum_{i=1}^{k} y_i\right),$$

cuja média e variância são, respectivamente:

$$E(\pi|\boldsymbol{y}) = \frac{\alpha + \sum_{i=1}^{k} y_i}{\alpha + \beta + \sum_{i=1}^{k} n_i} = \mu_\pi, \tag{7}$$

$$V(\pi|\boldsymbol{y}) = \frac{\left(\alpha + \sum_{i=1}^{k} y_i\right)\left(\beta + \sum_{i=1}^{k} n_i - \sum_{i=1}^{k} y_i\right)}{\left(\alpha + \beta + \sum_{i=1}^{k} n_i\right)^2 \left(\alpha + \beta + \sum_{i=1}^{k} n_i + 1\right)} = \sigma_\pi^2. \tag{8}$$

A formalização do teste bayesiano (TB) foi inspirada nas estatísticas multivariadas de razão de verossimilhanças para vetores de médias normais, que é uma forma quadrática (Johnson e Weerahandi, 1988; Ferreira, 2008). Assim, para realizar o teste será considerada a função dos parâmetros dada por:

$$q_c = (\boldsymbol{\pi}_p - \boldsymbol{\pi}_0)^\top \boldsymbol{\Sigma}_p^{-1}(\boldsymbol{\pi}_p - \boldsymbol{\pi}_0), \tag{9}$$

em que $\boldsymbol{\pi}_p$ é a média da distribuição *a posteriori* de $\boldsymbol{\pi}$, dada por:

$$\boldsymbol{\pi}_p = \begin{bmatrix} E(\pi_1|\boldsymbol{y}) \\ E(\pi_2|\boldsymbol{y}) \\ \vdots \\ E(\pi_k|\boldsymbol{y}) \end{bmatrix} = E(\boldsymbol{\pi}|\boldsymbol{y}),$$

tendo $E(\pi_i|\boldsymbol{y})$ sido apresentada em (3), $\boldsymbol{\pi}_0$ é um vetor de médias de k cadeias independentes de π, sob H_0, dado por:

$$\boldsymbol{\pi}_0 = \begin{bmatrix} \mu_\pi \\ \mu_\pi \\ \vdots \\ \mu_\pi \end{bmatrix}, \tag{10}$$

em que a média μ_π é dada em (7) e $\boldsymbol{\Sigma}_p^{-1}$ é a inversa da matriz de covariâncias da distribuição *a posteriori*, sob H_1, de $\boldsymbol{\pi}$, que é dada por:

$$\boldsymbol{\Sigma}_p = \begin{bmatrix} \sigma^2_{\pi_1} & 0 & \dots & 0 \\ 0 & \sigma^2_{\pi_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma^2_{\pi_k} \end{bmatrix} = \mathrm{diag}(\sigma^2_{\pi_i}),$$

em que $\sigma^2_{\pi_i}$ é dado em (4).

Como a distribuição da função paramétrica (9) é desconhecida, usou-se o procedimento descrito na sequência. Inicialmente, foram geradas k cadeias independentes da distribuição *a posteriori* de π sob H_0, emulando uma situação de k diferentes populações. Essa é a mesma lógica da distribuição da amplitude estudentizada para comparações múltiplas de k médias normais (Steel e Torrie, 1980; Hinkelmann e Kempthorne, 1987). Assim, foram gerados vetores $\boldsymbol{\pi_j}$, cujos componentes π_{ij} são realizações da distribuição *a posteriori* beta:

$$\pi_{ij}|\boldsymbol{y} \sim B\left(\alpha + \sum_{i=1}^{k} y_i, \beta + \sum_{i=1}^{k} n_i - \sum_{i=1}^{k} y_i\right),$$

em que π_{ij} é a realização aleatória da distribuição beta referente a i-ésima população (emulação) e j-ésima unidade da distribuição de equilíbrio *a posteriori*, sendo $j = 1, 2, \cdots, N$ e $i = 1, 2, \cdots, k$. Foi considerado um valor de N igual a 10.000. É relevante salientar que as distribuições *a posteriori*, sob H_1 ou sob H_0, são conhecidas, o que possibilita a obtenção de amostras válidas diretamente por meio de simulação Monte Carlo, sem a necessidade de utilização de cadeias de Markov. O vetor de médias das k distribuições *a posteriori* sob H_0 foi apresentado em (10) e a matriz de covariâncias $\boldsymbol{\Sigma}_0$ é dada por:

$$\boldsymbol{\Sigma}_0 = \begin{bmatrix} \sigma^2_{\pi} & 0 & \dots & 0 \\ 0 & \sigma^2_{\pi} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma^2_{\pi} \end{bmatrix} = \mathrm{diag}(\sigma^2_{\pi}),$$

em que σ^2_π é dado em (8).

Para cada unidade vetorial j da distribuição de equilíbrio *a posteriori* sob H_0 simulada ($\boldsymbol{\pi}_j$), foi obtida a forma quadrática:

$$q_j = (\boldsymbol{\pi}_j - \boldsymbol{\pi}_0)^\top \Sigma_0^{-1} (\boldsymbol{\pi}_j - \boldsymbol{\pi}_0), \tag{11}$$

em que $\boldsymbol{\pi_j} = [\ \pi_{1j}, \quad \pi_{2j}, \quad \ldots, \quad \pi_{kj}\ ]^\top$, para $j = 1, 2, \cdots, N$.

A quantidade (11) pode ser simplificada por:

$$q_j = \sum_{i=1}^{k} (\pi_{ij} - \mu_\pi)^2 \left[\frac{\left(\alpha + \beta + \sum_{i=1}^{k} n_i\right)^2 \left(\alpha + \beta + \sum_{i=1}^{k} n_i + 1\right)}{\left(\alpha + \sum_{i=1}^{k} y_i\right) \left(\beta + \sum_{i=1}^{k} n_i - \sum_{i=1}^{k} y_i\right)} \right]. \tag{12}$$

Os N valores de q_j, $j = 1, 2, \cdots, N$ obtidos formaram a distribuição nula da estatística teste. Os valores foram ordenados e o quantil superior $100\alpha\%$ q_α foi obtido para $\alpha = 5\%$ e 1%. A decisão, contrária ou não à hipótese H_0, foi tomada de acordo com o seguinte critério: o valor da função paramétrica q_c de (9) foi confrontado com esse quantil e, quando $q_c > q_\alpha$, a hipótese nula foi rejeitada ou computando o valor de credibilidade (valor-c) a favor de H_0 por:

$$\text{valor} - c = \frac{\displaystyle\sum_{j=1}^{N} I(q_j \geq q_c)}{N},$$

em que $I(q_j \geq q_c)$ é a função indicadora que deve retornar 1 se $q_j \geq q_c$ ou 0, em caso contrário. Se o valor-c for inferior a um valor de credibilidade nominal α, deve-se rejeitar a hipótese nula.

2.2 Comparações múltiplas bayesianas

As comparações múltiplas são definidas para todos os testes de hipóteses entre duas proporções binomiais. Assim, para testar H_0: $\boldsymbol{\ell}^\top \boldsymbol{\pi} = 0$, em que o vetor $\boldsymbol{\ell}$ possui na i-ésima e i'-ésima posições os valores de 1 e -1, respectivamente, e nas demais o valor 0, foi utilizada a distribuição nula de q_j, definida em (12), propondo a seguinte função paramétrica:

$$\delta_c = \frac{\boldsymbol{\ell}^\top \boldsymbol{\pi}_p}{\sqrt{\boldsymbol{\ell}^\top \boldsymbol{\Sigma}_p^{-1} \boldsymbol{\ell}}}.$$

Essa quantidade pode ser expandida na seguinte expressão:

$$\delta_c = \frac{\left[\frac{\alpha_i + y_i}{\alpha_i + \beta_i + n_i} - \frac{\alpha_{i'} + y_{i'}}{\alpha_{i'} + \beta_{i'} + n_{i'}}\right]}{\sqrt{\left[\frac{(\alpha_i + \beta_i + n_i)^2 (\alpha_i + \beta_i + n_i + 1)}{(\alpha_i + y_i)(\beta_i + n_i - y_i)} + \frac{(\alpha_{i'} + \beta_{i'} + n_{i'})^2 (\alpha_{i'} + \beta_{i'} + n_{i'} + 1)}{(\alpha_{i'} + y_{i'})(\beta_{i'} + n_{i'} - y_{i'})}\right]}}. \tag{13}$$

Portanto, se $|\delta_c| \geq \sqrt{q_\alpha}$, a hipótese nula H_0: $\boldsymbol{\ell}^\top \boldsymbol{\pi} = 0$ deve ser rejeitada, considerando esse nível α de probabilidade nominal. É conveniente destacar que qualquer outra escolha do vetor $\boldsymbol{\ell}$ pode ser feita, embora o foco, neste trabalho, tenha sido o de comparações múltiplas. Nesse contexto, é possível obter o intervalo de credibilidade para $\boldsymbol{\ell}^\top \boldsymbol{\pi}$ da seguinte forma:

$$\mathrm{IC}_{1-\alpha}(\boldsymbol{\ell}^\top \boldsymbol{\pi}) : \boldsymbol{\ell}^\top \boldsymbol{\pi}_p \pm \sqrt{q_\alpha}\sqrt{\boldsymbol{\ell}^\top \boldsymbol{\Sigma}_p^{-1} \boldsymbol{\ell}} \tag{14}$$

e a região de credibilidade (RC) de $100(1-\alpha)\%$ para $\boldsymbol{\pi}$ por:

$$\mathrm{RC}_{1-\alpha}(\boldsymbol{\pi}) : \{\boldsymbol{\pi} | (\boldsymbol{\pi} - \boldsymbol{\pi_p})^\top \boldsymbol{\Sigma}_p^{-1} (\boldsymbol{\pi} - \boldsymbol{\pi_p}) \leq q_\alpha\}. \tag{15}$$

As expressões (14) e (15) não foram utilizadas diretamente na avaliação do teste de comparações múltiplas bayesiano (TCMB), mas possuem grande valor didático e científico e por isso foram apresentadas.

2.3 Simulação Monte Carlo

A avaliação do desempenho dos testes bayesianos propostos neste trabalho foi feita por meio de simulação Monte Carlo. Foram geradas k populações binomiais independentes, com parâmetros $\pi_1, \pi_2, \cdots, \pi_k$ e $n_1, n_2, \cdots, n_k$. Foram realizadas 1.000 simulações Monte Carlo para cada configuração, envolvendo combinações das quantidades k, n_i's e π_i's. As simulações foram subdivididas em duas partes, sendo a primeira para o TB e a segunda para as comparações múltiplas bayesianas. Cada uma delas foi subdividida novamente em duas etapas. Na primeira, foram simuladas situações sob H_0 completa para se avaliar o erro tipo I do teste bayesiano (TB) e o erro tipo I por experimento do teste de comparações múltiplas bayesiano (TCMB). Na segunda, foram simuladas situações sob H_0 parcial para avaliar o poder do TB e do TCMB e, ainda, as taxas de erro tipo I por experimento somente do TCMB.

Para a simulação sob H_0 completa foram consideradas populações com π_i's idênticos e fixados em 0,01, 0,1 e 0,5. Sob H_0 parcial, dois grupos iguais internamente, mas diferentes entre si, foram considerados. Nesse caso, a diferença (Δ) entre os valores dos parâmetros dos dois grupos, foram iguais a 0,01, 0,05, 0,1, 0,2, 0,3, 0,4, 0,5, 0,6, 0,7, 0,8 e 0,9. O valor do parâmetro π_ℓ do primeiro grupo foi fixado em 0,01, sendo $\ell = 1, 2, 3$ quando o número de populações binomiais foi igual a 5 ($k = 5$) e $\ell = 1, 2, \cdots, 5$ no caso de $k = 10$. Também foi considerada uma situação em que o valor π_ℓ do primeiro grupo foi fixado em 0,3, 0,45 e 0,5 e os valores de Δ em 0,01, 0,10 e 0,4.

Foram consideradas $k = 2, 5$ e 10 populações e tamanhos de amostras $n_i =$ 10, 30 e 100 para cada população, $\forall\ i = 1, 2, \cdots, k$. Foram fixados os níveis nominais de probabilidade $\alpha = 1\%$ e 5%. As taxas de erro tipo I e poder foram computadas para os testes em questão conforme a situação.

Foi realizado um teste binomial exato, considerando a hipótese $H_0 : \alpha = 5\%$ vs $H_1 : \alpha \neq 5\%$ e $H_0 : \alpha = 1\%$ vs $H_1 : \alpha \neq 1\%$, para um nível nominal de probabilidade de 1%. O teste será considerado conservativo nas situações em que a hipótese de nulidade (H_0) for rejeitada e a taxa de erro tipo I observada for considerada significativamente ($P < 0{,}01$) inferior ao nível nominal. Caso contrário, se a taxa de erro tipo I observada for considerada significativamente ($P < 0{,}01$) superior ao nível nominal, o teste será considerado liberal. Por fim, pode acontecer, ainda, de a taxa de erro tipo I observada não ser significativamente diferente do nível nominal ($P > 0{,}01$), situação ideal e, nesses casos, o teste será considerado exato.

A estatística do teste foi obtida da relação entre as distribuições binomial e F, com probabilidade de sucesso $\pi = \alpha$, considerando que m representa o número de hipóteses nulas rejeitadas em N simulações Monte Carlo para o nível nominal α. Essa estatística é dada por:

$$F = \left(\frac{m+1}{N-m}\right)\left(\frac{1-\alpha}{\alpha}\right),$$

que, sob a hipótese nula, segue uma distribuição F com $\nu_1 = 2(N-m)$ e $\nu_2 = 2(m+1)$ graus de liberdade. Quando for verificado que $F \leq F_{0,005}$ ou que $F \geq F_{0,995}$, a hipótese nula pode ser rejeitada ao nível nominal de probabilidade de 1%, em que $F_{0,005}$ e $F_{0,995}$ são os quantis da distribuição F com ν_1 e ν_2 graus de liberdade.

Os hiperparâmetros da distribuição *a priori* foram fixados por tentativa e erro, buscando minimizar as taxas de erro tipo I e maximizar o poder. Assim, inúmeras situações foram avaliadas quanto aos valores de α_i's, β_i's, α's e β's, sob H_0 completa e parcial. Os valores que trouxeram um controle adequado do erro tipo I e maiores valores de poder foram escolhidos, tornando o teste final mais sensível. Dessa forma, os valores dos hiperparâmetros para o TB sob H_1 foram ($\alpha_i = 2$, $\beta_i = 2$) e ($\alpha_i = 1$, $\beta_i = 1$), para $i = 1, 2, \cdots, k$ e, sob H_0, foram ($\alpha = 2$, $\beta = 2$) e ($\alpha = 1$, $\beta = 1$). Para as comparações múltiplas, todos os hiperparâmetros foram fixados em 2 e 0,01, para os α_i's e β_i's e para os α's e β's, respectivamente.

Os resultados do TB foram comparados com os obtidos para os testes de razão de verossimilhanças e X^2 de Pearson, baseado em formas quadráticas, denotados, respectivamente, por G^2 e X^2, apresentados por Biase e Ferreira (2009). O desempenho do TCMB também foi confrontado com o obtido para os testes de *bootstrap*, avaliados por Biase e Ferreira (2006) e com o teste de comparações múltiplas assintóticas, baseado em formas quadráticas, proposto por Biase e Ferreira (2009).

3 Resultados e discussão

As simulações foram realizadas considerando os níveis nominais de probabilidade de 1% e 5%. No entanto, devido a grande similaridade em relação aos resultados para as taxas de erro tipo I e poder dos testes para $\alpha = 1\%$ e 5%, foram apresentados apenas os resultados para $\alpha = 5\%$.

Para avaliar o desempenho do teste bayesiano, as taxas de erro tipo I e poder foram computadas e são apresentadas e discutidas separadamente nas subseções 3.1 e 3.2.

3.1 Erro tipo I sob H_0 completa

As taxas de erro tipo I, em porcentagem, sob H_0 completa do teste bayesiano, representado por TB, são apresentadas na Tabela 1 em função de k, n e π, considerando hiperparâmetros iguais a 2 (α_i's, β_i's, α's e β's) e nível nominal de 5%. Todos os resultados correspondem a médias de 1.000 simulações Monte Carlo.

Tabela 1 - Taxas de erro tipo I (%), sob H_0 completa, para diferentes números de populações (k), de tamanhos de amostras (n) e de valores do parâmetro (π), para o teste bayesiano (TB) com todos os hiperparâmetros iguais a 2, ao nível nominal de 5%

		π		
k	n	0,01	0,1	0,5
2	10	$0{,}00^+$	$0{,}00^+$	$0{,}70^+$
2	30	$0{,}00^+$	$0{,}40^+$	$1{,}40^+$
2	100	$0{,}00^+$	$1{,}60^+$	$1{,}30^+$
5	10	$0{,}00^+$	$0{,}00^+$	$1{,}90^+$
5	30	$0{,}00^+$	$0{,}00^+$	$2{,}50^+$
5	100	$0{,}00^+$	$1{,}10^+$	3,60
10	10	$0{,}00^+$	$0{,}00^+$	$3{,}10^+$
10	30	$0{,}00^+$	$0{,}00^+$	$3{,}00^+$
10	100	$0{,}00^+$	$2{,}10^+$	$3{,}20^+$

$^+$ significativamente inferior ao nível nominal de 5%, considerando uma credibilidade de 99%.

De modo geral, pode-se observar que houve controle do erro tipo I, pois não ocorreram situações em que as taxas de erro tipo I superaram significativamente ($P < 0{,}01$) o valor nominal de 5%. O que se verifica, na grande maioria dos casos, é que o TB foi considerado conservativo, ou seja, as taxas de erro tipo I foram significativamente inferiores ($P < 0{,}01$), a 5%. Observa-se também que, para valores de π próximos de 0,5 e para número de populações maiores ($k = 5$ e 10), houve uma tendência de as taxas de erro tipo I aproximarem-se do valor nominal adotado.

A única situação em que o TB apresentou taxa de erro tipo I igual ao valor nominal foi considerando $\pi = 0{,}5$, com $k = 5$ e $n = 100$. Nas demais situações, o tamanho do TB foi significativamente inferior ao valor nominal de 5%. Em nenhuma das configurações avaliadas, o TB apresentou desempenho liberal, ou seja, taxas de erro tipo I significativamente ($P < 0{,}01$) superiores ao nível nominal. Isso mostra que esse teste controlou o erro tipo I sob a hipótese H_0 completa, embora de forma conservativa.

As taxas de erro tipo I sob H_0 completa do TB também foram avaliadas considerando os valores dos hiperparâmetros, α_i's e β_i's iguais a 1 e dos α's e β's iguais a 2. Na Tabela 2 são apresentadas essas taxas em função de k, n e π, para $\alpha = 5\%$. Em várias situações, o que se observa, de modo geral, é que as taxas de erro tipo I tenderam a se aproximar do nível nominal de 5%, passando a existir casos em que as taxas foram consideradas significativamente ($P < 0{,}01$) superiores a 5%.

Tabela 2 - Taxas de erro tipo I (%), sob H_0 completa, para diferentes números de populações (k), tamanhos de amostras (n) e diferentes valores do parâmetro (π) para o teste bayesiano (TB) com hiperparâmetros α_i's e β_i's iguais a 1 e α's e β's iguais a 2, ao nível nominal de 5%

		π		
k	n	0,01	0,1	0,5
2	10	$0{,}00^+$	$1{,}60^+$	$1{,}20^+$
2	30	$0{,}00^+$	$1{,}90^+$	$1{,}70^+$
2	100	$0{,}10^+$	$1{,}30^+$	$1{,}60^+$
5	10	$0{,}00^+$	$0{,}30^+$	5,50
5	30	$0{,}00^+$	$3{,}10^+$	3,50
5	100	$0{,}00^+$	4,20	3,60
10	10	$0{,}00^+$	$0{,}10^+$	$8{,}20^*$
10	30	$0{,}00^+$	$2{,}50^+$	5,00
10	100	$0{,}00^+$	4,50	3,30

* significativamente superior ao nível nominal de 5%, considerando uma credibilidade de 99%.
$^+$ significativamente inferior ao nível nominal de 5%, considerando uma credibilidade de 99%.

O tamanho do TB não foi significativamente diferente do valor nominal de 5% com $k = 5$ e 10 e $n = 100$, para $\pi = 0{,}1$. Para $\pi = 0{,}5$, verificou-se que, para $k = 5$ e 10, o TB foi considerado exato, com exceção para $k = 10$ e $n = 10$, em que a taxa de erro tipo I do teste foi significativamente ($P < 0{,}01$) superior, a 5%, sendo, nesse caso, o teste classificado como liberal. Nas demais situações, as taxas de erro tipo I foram significativamente ($P < 0{,}01$) inferiores a 5%, indicando desempenho conservativo do TB.

Com base nesses resultados, observa-se que o TB apresentou melhor desempenho, considerando todos os hiperparâmetros iguais a 2, uma vez que o teste em questão não foi considerado liberal em nenhuma situação avaliada, se comparado aos resultados obtidos das taxas de erro tipo I, considerando os hiperparâmetros iguais a 1 e 2. Agresti e Coull (1998), avaliando métodos de estimação intervalar das proporções binomiais, verificaram que, substituindo-se o estimador de máxima verossimilhança pelo estimador das proporções add-4 no intervalo de Wald, os resultados obtidos foram surpreendentes, uma vez que passaram de extremamente liberais para expressivamente conservativos. Segundo esses autores, o ponto médio desse intervalo é dado pela estimativa pontual desse estimador, que corresponde também à estimativa de Bayes (média da distribuição *a posteriori*) considerando uma distribuição *a priori* beta com parâmetros (α e β) iguais a 2, tendo média de 0,5 e desvio padrão de 0,224. Isso torna-se propício à realização do TB utilizando-se hiperparâmetros iguais a 2.

Verificou-se também que, o TB é muito mais conservativo do que o teste X^2 e, principalmente em relação ao teste G^2. Esses testes X^2 e G^2 foram avaliados por Biase e Ferreira (2009). Provavelmente, o poder do TB será inferior ao poder dos testes G^2 e X^2.

3.2 Poder sob H_0 parcial

Na Tabela 3 é apresentado, em porcentagem, o poder sob H_0 parcial do TB em função de k, n e Δ, considerando hiperparâmetros iguais a 2 e $\alpha = 5\%$. Nessas situações, consideraram-se valores de π no primeiro grupo iguais a 0,01 e, no segundo, iguais a $0{,}01 + \Delta$. Se os valores de Δ são muito pequenos ($\Delta = 0{,}01$ ou 0,05), o poder do TB é igual a zero ou pequeno (inferior a 70%), independente do número de populações (k) e do tamanho amostral (n).

Tabela 3 - Poder (%), sob H_0 parcial, para diferentes números de populações (k), de tamanhos de amostras (n) e de diferenças entre os parâmetros binomiais π de cada grupo (Δ), para o teste bayesiano (TB) com todos os hiperparâmetros iguais a 2, ao nível nominal de 5%

		Δ								
k	n	0,01	0,05	0,1	0,2	0,3	0,4	0,5	0,7	0,9
5	10	0,00	0,00	0,00	0,90	6,90	29,90	62,00	98,60	100,00
5	30	0,00	0,00	0,90	42,60	92,30	99,90	100,00	100,00	100,00
5	100	0,00	9,90	87,60	100,00	100,00	100,00	100,00	100,00	100,00
10	10	0,00	0,00	0,00	2,00	19,30	64,30	95,70	100,00	100,00
10	30	0,00	0,00	1,10	86,20	99,90	100,00	100,00	100,00	100,00
10	100	0,00	30,00	99,80	100,00	100,00	100,00	100,00	100,00	100,00

Para diferenças pequenas ou moderadas ($0{,}1 \leq \Delta \leq 0{,}3$), observa-se que o poder do TB é pequeno em todos os casos, considerando pequenas amostras ($n = 10$) e, em algumas situações, considerando amostras intermediárias ($n = 30$). Para grandes amostras ($n = 100$), o poder do teste tende a se aproximar de 100%. Entretanto, para valores de Δ grandes ou muito grandes, $\Delta \geq 0{,}4$, o poder do TB aproximou-se rapidamente de 100%, na maioria das situações avaliadas, inclusive para pequenas amostras. Isso não foi verificado apenas com $k = 5$ e $n = 10$, considerando $\Delta = 0{,}4$ e 0,5 e também com $k = 10$ e $n = 10$, para $\Delta = 0{,}4$.

De maneira geral, pode-se observar que houve um grande efeito do tamanho das amostras (n) e do número de populações (k), no sentido de aumentar o poder do TB. Assim, ao se fixar um valor de Δ e o número de populações (k), o aumento de n proporciona acréscimos consideráveis do poder, principalmente se a diferença Δ é pequena ou moderada. O mesmo desempenho do poder do TB é observado se forem fixados um Δ e o tamanho amostral (n), variando o número de populações de 5 para 10.

O poder do TB também foi avaliado sob a hipótese H_0 parcial, considerando os hiperparâmetros da distribuição *a priori* α_i's e β_i's iguais a 1 e α's e β's iguais a 2. Na Tabela 4 é apresentado esse poder, em função de k, n e Δ, para $\alpha = 5\%$. De modo geral, verifica-se que o desempenho do teste, nesse caso, apresentou semelhanças ao observado para o poder do TB considerando todos os hiperparâmetros iguais a 2. No entanto, observa-se que, na maioria das situações, o poder do TB com hiperparâmetros iguais a 1 e 2 foi superior ao poder do TB, considerando apenas hiperparâmetros iguais a 2.

Tabela 4 - Poder (%), sob H_0 parcial, para diferentes números de populações (k), de tamanhos de amostras (n) e de diferenças entre os parâmetros binomiais π de cada grupo (Δ), para o teste bayesiano (TB) com hiperparâmetros α_i's e β_i's iguais a 1 e α's e β's iguais a 2, ao nível nominal de 5%

		Δ								
k	n	0,01	0,05	0,1	0,2	0,3	0,4	0,5	0,7	0,9
5	10	0,00	0,00	0,10	2,50	20,60	57,00	82,50	99,80	100,00
5	30	0,00	0,10	7,90	77,00	99,50	100,00	100,00	100,00	100,00
5	100	0,00	45,00	97,30	100,00	100,00	100,00	100,00	100,00	100,00
10	10	0,00	0,00	0,00	4,20	41,90	85,60	99,20	100,00	100,00
10	30	0,00	0,00	15,70	98,00	100,00	100,00	100,00	100,00	100,00
10	100	0,10	77,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00

Comparando-se esses resultados, pode-se observar que, para $\Delta = 0{,}01$, o poder do TB é relativamente igual para as duas configurações de hiperparâmetros considerados, independente do número de populações (k) e do tamanho amostral (n). Para valores de Δ maiores ($\Delta = 0{,}05$), o poder do TB, considerando os hiperparâmetros iguais a 1 e 2, é expressivamente superior ao poder do TB com hiperparâmetros iguais a 2, mas, somente para grandes amostras ($n = 100$).

Para valores de Δ pequenos ou moderados ($0{,}1 \leq \Delta \leq 0{,}3$), o poder do TB com hiperparâmetros iguais a 1 e 2 também apresentou várias situações em que o poder foi consideravelmente superior ao do TB, com hiperparâmetros iguais a 2. Para Δ grande ou muito grande ($\Delta \geq 0{,}4$), houve casos de superioridade do poder com hiperparâmetros iguais a 1 e 2, apenas em algumas situações de pequenas amostras ($n = 10$), uma vez que, para amostras intermediárias e grandes ($n \geq 30$), o poder do TB aproximou-se de 100%, independente dos valores dos hiperparâmetros.

Verificou-se também que, independente dos valores dos hiperparâmetros considerados para o TB, o poder desse teste é inferior ao poder dos testes G^2 e X^2, avaliado por Biase e Ferreira (2009), principalmente, nas situações de amostras pequenas e intermediárias ($n = 10$ e 30) e, diferenças de Δ pequenas e moderadas ($\Delta \leq 0{,}3$). A medida que os valores de Δ aumetam, a diferença entre o poder destes testes diminui, pois o poder de todos eles tende a se aproximar de 100%. A inferioridade do poder do TB era esperada, devido as taxas de erro tipo I desse teste serem mais conservativas do que as dos testes assintóticos apresentados por Biase e Ferreira (2009).

Buscando avaliar situações em que os valores das proporções populacionais (π) dos dois grupos estivessem próximos de 0,5, foram realizadas algumas simulações adicionais, nas quais avaliou-se o poder sob H_0 parcial do TB.

Na Tabela 5 é apresentado o poder do TB, com hiperparâmetros iguais a 2 em função de k, $\pi^{(1)}$, n e Δ, para $\alpha = 5\%$. Pode-se observar, para $\Delta = 0{,}01$, que o poder do TB é inferior ao nível nominal de significância de 5% para todos os tamanhos amostrais (n), número de populações (k) e valores de $\pi^{(1)}$.

Considerando $\Delta = 0{,}1$, o poder ainda é pequeno, no entanto, nota-se que o poder do TB aumenta à medida que o número de populações e tamanhos amostrais

Tabela 5 - Poder (%), sob H_0 parcial, para diferentes números de populações (k), de valores do parâmetro π no primeiro grupo ($\pi^{(1)}$), de tamanhos de amostras (n) e diferenças entre os parâmetros π de cada grupo (Δ), para o teste bayesiano (TB), com todos os hiperparâmetros iguais a 2, ao nível nominal de 5%

			Δ		
k	$\pi^{(1)}$	n	0,01	0,1	0,4
5	0,30	10	1,40	2,20	42,30
5	0,30	30	2,80	7,30	98,00
5	0,30	100	4,00	31,70	100,00
10	0,30	10	1,10	2,00	68,30
10	0,30	30	3,80	14,00	100,00
10	0,30	100	3,60	53,80	100,00
5	0,45	10	2,30	3,80	46,80
5	0,45	30	1,80	7,40	99,20
5	0,45	100	3,00	27,50	100,00
10	0,45	10	2,70	5,50	73,10
10	0,45	30	3,10	13,00	100,00
10	0,45	100	4,60	51,40	100,00
5	0,50	10	2,70	3,80	44,20
5	0,50	30	3,00	8,90	99,70
5	0,50	100	2,80	28,00	100,00
10	0,50	10	2,60	5,40	69,20
10	0,50	30	3,90	13,10	100,00
10	0,50	100	4,50	52,20	100,00

aumenta, principalmente se n aumenta de 30 para 100. Para uma diferença maior entre os valores de π dos grupos ($\Delta = 0{,}4$), verifica-se que, para $k = 5$ e $n = 10$, independente do valor de $\pi^{(1)}$ e também com $k = 10$, $n = 10$ e $\pi^{(1)} = 0{,}30$ e 0,50, o poder do TB foi pequeno (inferior a 70%). Nas demais situações, o poder desse teste tende a se aproximar de 100%.

Nesta situação em que o poder do TB foi avaliado considerando os valores de π do primeiro e segundo grupo próximos de 0,5, verificou-se, novamente que, o poder do TB foi inferior ao poder dos testes G^2 e X^2, apresentado por Biase e Ferreira (2009), considerando $\Delta = 0{,}01$ e 0,1. Para $\Delta = 0{,}4$ o desempenho do TB, G^2 e X^2 se igualou e aproximou-se de 100%, principalmente para grandes amostras ($n \geq 30$).

3.3 Teste de comparações múltiplas bayesiano

O teste de comparações múltiplas envolvendo proporções binomiais, utilizando-se a abordagem bayesiana, foi avaliado, inicialmente, em duas maneiras distintas de computar o erro tipo I por experimento: uma sob a hipótese H_0 completa e a outra sob a hipótese H_0 parcial. Posteriormente, o poder desse teste foi mensurado sob a hipótese H_0 parcial.

3.4 Erro tipo I

3.4.1 Erro tipo I sob H_0 completa

Na Tabela 6 são apresentadas, em porcentagem, as taxas de erro tipo I, por experimento, sob H_0 completa, do teste de comparações múltiplas bayesiano, identificado por TCMB, em função de k, n e π, considerando os hiperparâmetros da distribuição *a priori* beta iguais a 2, para $\alpha = 5\%$. Pode-se observar que houve controle do erro tipo I por experimento, pois, em todos os casos, as taxas de erro não superaram significativamente ($P < 0{,}01$) o valor nominal de 5%. Na verdade, observou-se que, exceto com $k = 5$, $n = 100$ e $\pi = 0{,}5$, em que o teste foi considerado exato, as taxas de erro tipo I por experimento do TCMB foram significativamente ($P < 0{,}01$) inferiores, ao nível nominal de 5%, o que indica que o TCMB é conservativo para essas situações avaliadas.

Tabela 6 - Taxas de erro tipo I, por experimento (%), sob H_0 completa, para diferentes números de populações (k), de tamanhos de amostras (n) e de valores do parâmetro π, para o teste de comparações múltiplas bayesiano (TCMB), com todos os hiperparâmetros iguais a 2, ao nível nominal de 5%

		π		
k	n	0,01	0,1	0,5
2	10	$0{,}00^+$	$0{,}00^+$	$1{,}50^+$
2	30	$0{,}00^+$	$0{,}40^+$	$2{,}00^+$
2	100	$0{,}00^+$	$0{,}40^+$	$0{,}60^+$
5	10	$0{,}00^+$	$0{,}00^+$	$0{,}30^+$
5	30	$0{,}00^+$	$0{,}00^+$	$1{,}30^+$
5	100	$0{,}00^+$	$0{,}20^+$	0,60
10	10	$0{,}00^+$	$0{,}00^+$	$0{,}00^+$
10	30	$0{,}00^+$	$0{,}00^+$	$0{,}10^+$
10	100	$0{,}00^+$	$0{,}00^+$	$0{,}00^+$

$^+$ significativamente inferior ao nível nominal de 5%, considerando uma credibilidade de 99%.

Observou-se que, o TCMB é extremamente conservativo em relação aos testes de *bootstrap* apresentados por Biase e Ferreira (2006) e ao teste de comparações múltiplas assintóticas, baseado em formas quadráticas, avaliado por Biase e Ferreira(2009), independente do número de populações (k), tamanhos amostrais (n) e valores dos parâmetros binomiais (π).

Procurando obter um melhor desempenho do TCMB, ou seja, taxas de erro tipo I por experimento iguais ao valor nominal de probabilidade, foram realizadas algumas simulações adicionais, variando-se os valores dos hiperparâmetros (α e β) da distribuição *a priori* do parâmetro π. Na Tabela 7, são apresentadas as taxas de erro tipo I por experimento do TCMB, considerando hiperparâmetros α_i's e β_i's iguais a 0,01 e α's e β's iguais a 2, em função de k, n e π, para o nível nominal de 5%. Verifica-se que, mesmo para pequenos valores dos hiperparâmetros, as taxas de erro para valores de π afastados de 0,5 ($\pi \leq 0{,}1$) são significativamente ($P < 0{,}01$) menores do que 5%. Portanto, o TCMB continuou sendo conservativo nesses casos.

Tabela 7 - Taxas de erro tipo I, por experimento (%), sob H_0 completa, para diferentes números de populações (k), de tamanhos de amostras (n) e de valores do parâmetro π para o teste de comparações múltiplas bayesiano (TCMB), com hiperparâmetros α_i's e β_i's iguais a 0,01 e α's e β's iguais a 2, ao nível nominal de 5%

		π		
k	n	0,01	0,1	0,5
2	10	$0{,}00^{+}$	$0{,}50^{+}$	4,60
2	30	$0{,}00^{+}$	$1{,}80^{+}$	3,90
2	100	$0{,}00^{+}$	$1{,}50^{+}$	$1{,}80^{+}$
5	10	$0{,}00^{+}$	$0{,}10^{+}$	$10{,}00^{*}$
5	30	$0{,}00^{+}$	$0{,}60^{+}$	$2{,}30^{+}$
5	100	$0{,}00^{+}$	$1{,}00^{+}$	$1{,}00^{+}$
10	10	$0{,}00^{+}$	$0{,}00^{+}$	$8{,}90^{*}$
10	30	$0{,}00^{+}$	$0{,}00^{+}$	$0{,}60^{+}$
10	100	$0{,}00^{+}$	$0{,}20^{+}$	$0{,}30^{+}$

* significativamente superior ao nível nominal de 5%, considerando uma credibilidade de 99%.
$^{+}$ significativamente inferior ao nível nominal de 5%, considerando uma credibilidade de 99%.

Para valores de $\pi = 0{,}5$, o TCMB apresentou melhores resultados para $k = 2$ e tamanhos de amostras pequenas e intermediárias ($n = 10$ e 30), em que a taxa não diferiu significativamente ($P > 0{,}01$) do valor nominal adotado. Houve também alguns casos em que o TCMB foi liberal, isto é, apresentou taxas de erro tipo I significativamente ($P < 0{,}01$) superiores a 5%. Isso ocorreu para pequenas amostras ($n = 10$) e $k = 5$ e 10. Nas demais situações, o TCMB foi considerado conservativo.

3.4.2 Erro tipo I sob H_0 parcial

As taxas de erro tipo I por experimento, sob H_0 parcial, para o TCMB em função de k, n e Δ, são apresentadas na Tabela 8, considerando hiperparâmetros iguais a 2 e $\alpha = 5\%$. Verifica-se, de modo geral, que o TCMB controlou o erro tipo I por experimento sob H_0 parcial, independente do número de populações (k), do tamanho amostral (n) e da diferença entre os valores de π dos grupos (Δ). Em todos os casos, o TCMB foi conservativo, apresentando taxas relativamente próximas de zero, indicando excesso de conservadorismo.

O teste de comparações múltiplas assintóticas, baseado em formas quadráticas, proposto por Biase e Ferreira(2009), apresentou desempenho semelhante ao TCMB, sendo conservativo em todas as situações avaliadas. No entanto, as taxas de erro tipo I, por experimento, sob H_0 parcial do TCMB, foram expressivamente menores do que as do teste de comparações múltiplas assintóticas, baseado em formas quadráticas, o que pode ter efeito na redução do poder. Os testes de *bootstrap*

Tabela 8 - Taxas de erro tipo I, por experimento (%), sob H_0 parcial, para diferentes números de populações (k), de tamanhos de amostras (n) e diferenças entre os parâmetros π de cada grupo (Δ), para o teste de comparações múltiplas bayesiano (TCMB), com todos os hiperparâmetros iguais a 2, ao nível nominal de 5%

		Δ								
k	n	0,01	0,05	0,1	0,2	0,3	0,4	0,5	0,7	0,9
5	10	0,00$^+$	0,00$^+$	0,00$^+$	0,00$^+$	0,00$^+$	0,00$^+$	0,00$^+$	0,00$^+$	0,00$^+$
5	30	0,00$^+$	0,00$^+$	0,00$^+$	0,00$^+$	0,00$^+$	0,00$^+$	0,00$^+$	0,20$^+$	0,00$^+$
5	100	0,00$^+$	0,00$^+$	0,00$^+$	0,10$^+$	0,10$^+$	0,00$^+$	0,20$^+$	0,20$^+$	0,00$^+$
10	10	0,00$^+$	0,00$^+$	0,00$^+$	0,00$^+$	0,00$^+$	0,00$^+$	0,00$^+$	0,00$^+$	0,00$^+$
10	30	0,00$^+$	0,00$^+$	0,00$^+$	0,00$^+$	0,00$^+$	0,10$^+$	0,00$^+$	0,00$^+$	0,00$^+$
10	100	0,00$^+$	0,00$^+$	0,00$^+$	0,00$^+$	0,00$^+$	0,10$^+$	0,00$^+$	0,00$^+$	0,00$^+$

$^+$ significativamente inferior ao nível nominal de 5%, considerando uma credibilidade de 99%.

avaliados por Biase e Ferreira (2006) apresentaram melhor desempenho, pois, em alguns casos, o tamanho dos testes não foi significativamente ($P < 0{,}01$) diferente do valor nominal adotado.

Na Tabela 9 são apresentadas as taxas de erro tipo I por experimento, sob H_0 parcial do TCMB, com hiperparâmetros α_i's e β_i's iguais a 0,01 e α's e β's iguais a 2, em função de k, n e Δ, considerando o nível nominal de 5%. De maneira geral, verifica-se que os hiperparâmetros não tiveram grande influência nas taxas de erro tipo I sob H_0 parcial, pois, mesmo para hiperparâmetros pequenos, as taxas de erro do tipo I do TCMB apresentaram o mesmo padrão de resposta observado para as taxas de erro tipo I do TCMB com hiperparâmetros iguais a 2.

Tabela 9 - Taxas de erro tipo I, por experimento (%), sob H_0 parcial, para diferentes números de populações (k), de tamanhos de amostras (n) e diferenças entre os parâmetros π de cada grupo (Δ), para o teste de comparações múltiplas bayesiano (TCMB), com hiperparâmetros α_i's e β_i's iguais a 0,01 e α's e β's iguais a 2, ao nível nominal de 5%

		Δ								
k	n	0,01	0,05	0,1	0,2	0,3	0,4	0,5	0,7	0,9
5	10	0,00$^+$	0,00$^+$	0,00$^+$	0,10$^+$	0,90$^+$	0,80$^+$	1,10$^+$	0,60$^+$	0,00$^+$
5	30	0,00$^+$	0,00$^+$	0,00$^+$	0,00$^+$	0,20$^+$	0,20$^+$	0,40$^+$	0,50$^+$	0,00$^+$
5	100	0,00$^+$	0,00$^+$	0,00$^+$	0,20$^+$	0,30$^+$	0,10$^+$	0,10$^+$	0,30$^+$	0,10$^+$
10	10	0,00$^+$	0,10$^+$	0,00$^+$	0,10$^+$	1,00$^+$	2,10$^+$	2,30$^+$	0,80$^+$	0,00$^+$
10	30	0,00$^+$	0,00$^+$	0,00$^+$	0,10$^+$	0,10$^+$	0,20$^+$	0,10$^+$	0,10$^+$	0,00$^+$
10	100	0,00$^+$	0,00$^+$	0,20$^+$	0,10$^+$	0,00$^+$	0,00$^+$	0,00$^+$	0,10$^+$	0,10$^+$

$^+$ significativamente inferior ao nível nominal de 5%, considerando uma credibilidade de 99%.

Novamente, pode-se observar que todas as taxas de erro tipo I, por experimento sob H_0 parcial, foram significativamente ($P < 0{,}01$) inferiores ao nível nominal de 5%, indicando desempenho conservativo do TCMB. Porém, verifica-se que, para valores moderados e grandes de Δ ($0{,}3 \leq \Delta \leq 0{,}7$), algumas taxas de erro tipo I do TCMB, considerando hiperparâmetros iguais a 0,01 e 2, foram sutilmente superiores às observadas para o TCMB, considerando todos os hiperparâmetros iguais a 2.

Utilizando a aproximação bayesiana de Beal (1987) nos intervalos de confiança simultâneos para diferença entre proporções binomiais, Piegorsch (1991) também observou que as probabilidades de cobertura foram conservativas e, na maioria das situações, aproximaram do nível nominal adotado, considerando amostras pequenas e moderadas.

Para avaliar o TCMB com hiperparâmetros iguais a 2, considerando valores de π dos dois grupos próximos de 0,5, as taxas de erro tipo I por experimento, sob H_0 parcial, foram calculadas e são apresentadas na Tabela 10, em função de k, $\pi^{(1)}$ e n.

Tabela 10 - Taxas de erro tipo I, por experimento (%), sob H_0 parcial, para diferentes números de populações (k), valores do parâmetro π no primeiro grupo ($\pi^{(1)}$), tamanhos de amostras (n) e diferenças entre os parâmetros π de cada grupo (Δ), para o teste de comparações múltiplas bayesiano (TCMB), com todos os hiperparâmetros iguais a 2, ao nível nominal de 5%

			Δ		
k	$\pi^{(1)}$	n	0,01	0,1	0,4
5	0,30	10	0,00$^+$	0,30$^+$	0,10$^+$
5	0,30	30	0,40$^+$	0,40$^+$	0,30$^+$
5	0,30	100	0,20$^+$	0,30$^+$	0,20$^+$
10	0,30	10	0,10$^+$	0,00$^+$	0,00$^+$
10	0,30	30	0,00$^+$	0,00$^+$	0,00$^+$
10	0,30	100	0,10$^+$	0,00$^+$	0,00$^+$
5	0,45	10	0,20$^+$	0,00$^+$	0,10$^+$
5	0,45	30	0,00$^+$	0,60$^+$	0,20$^+$
5	0,45	100	0,00$^+$	0,70$^+$	0,10$^+$
10	0,45	10	0,10$^+$	0,00$^+$	0,00$^+$
10	0,45	30	0,10$^+$	0,00$^+$	0,00$^+$
10	0,45	100	0,00$^+$	0,10$^+$	0,00$^+$
5	0,50	10	0,00$^+$	0,50$^+$	0,30$^+$
5	0,50	30	0,40$^+$	0,30$^+$	0,20$^+$
5	0,50	100	0,50$^+$	0,30$^+$	0,20$^+$
10	0,50	10	0,00$^+$	0,00$^+$	0,00$^+$
10	0,50	30	0,20$^+$	0,00$^+$	0,00$^+$
10	0,50	100	0,10$^+$	0,00$^+$	0,00$^+$

$^+$ significativamente inferior ao nível nominal de 5%, considerando uma credibilidade de 99%.

Pode-se observar, de modo geral, que, em todas as situações, os valores das taxas de erro tipo I foram significativamente ($P < 0{,}01$) menores do que 5%. Ademais, verificou-se que, para número de populações menores ($k = 5$), as taxas de erro tipo I foram relativamente superiores às observadas para $k = 10$, enquanto os tamanhos amostrais não apresentaram interferência nos resultados obtidos.

Apesar de o TCMB ser considerado conservativo para valores de π próximos de 0,5, esses resultados são menos conservativos do que quando os valores de $\pi^{(1)}$ eram afastados de 0,5 ($\pi^{(1)} = 0{,}01$) (Tabela 8) sob H_0 parcial, considerando $k = 5$. Para $k = 10$, tanto para $\pi^{(1)}$ próximos ou afastados de 0,5, as taxas se aproximaram de zero.

3.5 Poder

Poder sob H_0 parcial

Como o poder do TCMB, com todos os hiperparâmetros iguais a 2, foi, em geral, bem inferior à situação em que os hiperparâmetros eram iguais a 0,01 e 2 e, em ambos os casos, houve controle do erro tipo I por experimento, embora de forma conservativa, então, apenas os resultados do último caso foram apresentados. O poder do TCMB considerando os hiperparâmetros iguais a 0,01 e 2 em função de k, n e Δ, é apresentado na Tabela 11 para $\alpha = 5\%$.

Tabela 11 - Poder (%), sob H_0 parcial, para diferentes números de populações (k), de tamanhos de amostras (n) e diferenças entre os parâmetros π de cada grupo (Δ), para o teste de comparações múltiplas bayesiano (TCMB), com hiperparâmetros α_i's e β_i's iguais a 0,01 e α's e β's iguais a 2, ao nível nominal de 5%

		Δ								
k	n	0,01	0,05	0,1	0,2	0,3	0,4	0,5	0,7	0,9
5	10	0,00	0,00	0,00	0,85	6,40	17,95	39,78	87,03	99,83
5	30	0,00	0,03	0,63	16,46	59,60	88,80	98,50	100,00	100,00
5	100	0,00	1,76	35,30	97,20	100,00	100,00	100,00	100,00	100,00
10	10	0,00	0,02	0,00	0,12	1,35	6,18	18,04	64,83	98,78
10	30	0,00	0,00	0,00	1,02	15,01	53,99	87,05	99,91	100,00
10	100	0,00	0,00	3,66	76,10	99,42	100,00	100,00	100,00	100,00

Pode-se observar que, para valores muito pequenos de Δ ($\Delta = 0{,}01$ e 0,05), o poder do TCMB é pequeno, apresentando, na maioria das situações, valores aproximadamente iguais a zero. Se a diferença Δ é pequena ou moderada, $0{,}1 \leq \Delta \leq 0{,}3$, o poder tende a se aproximar de 100% somente para grandes amostras ($n = 100$), exceto para $k = 5$ e 10, com $n = 100$, considerando $\Delta = 0{,}1$.

Para grandes valores de Δ ($\Delta = 0{,}4$ e 0,5), verifica-se que o poder do TCMB aumenta expressivamente com o aumento do tamanho das amostras (n), principalmente se (n) aumenta de 10 para 30. Se as diferenças são muito grandes ($\Delta \geq 0{,}7$), o poder é pequeno apenas para $k = 10$, $n = 10$, considerando $\Delta = 0{,}7$. Nos demais casos, o poder do TCMB tende a se aproximar de 100%, com o aumento das amostras.

De modo geral, verifica-se que o aumento do tamanho das amostras exerce influência favorável no crescimento do poder do teste, ao contrário do que ocorre com o aumento do número de populações (k), que propicia a redução do poder. Essas informações estão de acordo com as obtidas para o poder do teste de comparações múltiplas assintóticas, baseado em formas quadráticas, apresentado por Biase e Ferreira (2009) e, também com os testes de *bootstrap* avaliados por Biase e Ferreira (2006). Entretanto, observou-se que, para $\Delta \leq 0{,}4$, o TCMB apresentou poder inferior ao teste de comparações múltiplas assintóticas, baseado em formas quadráticas, na maioria das situações avaliadas. Mas, à medida que a diferença Δ tornou-se maior ($\Delta \geq 0{,}5$), o poder do TCMB aproximou-se de 100%, na maioria dos casos, e igualou-se ao poder do teste de comparações múltiplas assintóticas, baseado em formas quadráticas.

Conclusões

O teste bayesiano para a igualdade de várias proporções binomiais apresentou excelente desempenho, controlando o erro tipo I em praticamente todas as situações, em níveis iguais ou inferiores aos valores nominais. O poder desse teste é relativamente alto, principalmente se as diferenças entre as proporções binomiais dos dois grupos são grandes.

O teste de comparações múltiplas para proporções binomiais foi proposto com sucesso. De maneira geral, o teste é conservativo, sob H_0 completa e parcial e apresenta valores de poder altos.

BIASE, N. G.; FERREIRA, D. F. Tests of equality and multiple comparisons of several independent binomial proportions. *Rev. Bras. Biom.*, São Paulo, v.29, n.4, p.549-570, 2011.

▪ *ABSTRACT: In frequentist theory, inferences about more than two binomial proportions are performed using analysis of variance and multiple comparisons procedures, or still, by means of asymptotic tests and computational intensive methods. All these procedures have limitations due to violations of some of the assumptions required by tests. Therefore, this work aimed to propose a bayesian multiple comparisons test for proportions and a binomial test for the equality of several binomial proportions, and also to evaluate their performance using Monte Carlo simulation. Independent binomial populations with parameters* π_i *and* n_i, $i = 1, 2, \cdots, k$ *were considered and Monte Carlo simulations were performed for each configuration involving combinations the quantities* k, n_i*'s and* π_i*'s, considering conjugate betas prior with parameters* α_i *and* β_i*, settled by trial and error for minimizing the type I error rates and maximizing the power. The TB for equality of several binomial proportions showed excellent performance and relatively high power. The TCMB for binomial proportions, under complete and partial* H_0*, was conservative and showed high power.*

▪ *KEYWORDS: Monte Carlo simulation; bayesian test for equality of several binomial proportions; bayesian multiple comparisons test; type I error; power.*

Referências

AGRESTI, A.; COULL, B. A. Approximate is better than “exact” for interval estimation of binomial proportions. *Am. Stat.*, Alexandria, v.52, n.2, p.119-126, 1998.

AGRESTI, A.; MIN, Y. Frequentist performance of bayesian confidence intervals for comparing proportions in 2 x 2 contingency tables. *Biometrics*, Washington, v.61, n.2, p.515-523, 2005.

ANDRADE, P. C. R. *Comparações múltiplas bayesianas em modelos normais homocedásticos e heterocedásticos*. 2008. 96f. Tese (Doutorado em Estatística) - Universidade Federal de Lavras, Lavras, 2008.

BEAL, S. L. Asymptotic confidence intervals for the difference between two binomial parameters for use with small samples. *Biometrics*, Washington, v.43, n.4, p.941-950, 1987.

BIASE, N. G.; FERREIRA, D. F. Comparações múltiplas e testes simultâneas para parâmetros binomiais de k populações independentes. *Rev. Bras. Biom.*, São Paulo, v.27, n.3, p.301-323, 2009.

BIASE, N. G.; FERREIRA, D. F. Comparações múltiplas para proporções binomiais utilizando bootstrap. *Rev. Bras. Biom.*, São Paulo, v.24, n.1, p.95-112, 2006.

BRATCHER, T.; HAMILTON, C. A bayesian multiple comparison procedure for ranking the means of normally distributed data. *J. Stat. Plann. Infer.*, Amsterdam, v.133, n.1, p.23-32, 2005.

DOBSON, A. J.; BARNETT, A. *An introduction to generalized linear models*. 3.ed. London: Chapman and Hall, 2008. 244p.

FERREIRA, D. F. *Estatística multivariada*. v.1. Lavras: UFLA, 2008. 662p.

HINKELMANN, K.; KEMPTHORNE, O. *Design and analysis of experiments*, New York: J. Wiley, 1987. 445p. v.1.

JOHNSON, R. A.; WEERAHANDI, S. A bayesian solution to the multivariate Behrens-Fisher problem. *J. Am. Stat. Assoc.*, New York, v.83, n.401, p.145-149, 1988.

KRISHNAMOORTHY, K.; PENG, J. Exact properties of a new test and other tests for differences between several binomial proportions. *J. Appl. Stat. Sci.*, New York, v.16, n.4, p.23-35, 2008.

KRISHNAMOORTHY, K.; THOMSON, J.; CAI, Y. An exact method of testing equality of several binomial proportions to a specified standard. *Comput. Stat. Data Anal.*, Amsterdam, v.45, p.697-707, 2004.

MCCULLOCH, C. E.; SEARLE, S. R. *Generalized, linear, and mixed models*. New York: J. Wiley, 2001. 346p.

NELDER, J. A.; WEDDERBURN, R. W. M. Generalized linear models. *J. R. Stat. Soc. Ser. A: Stat. Soc.*, London, v.135, n.3, p.370-384, 1972.

PIEGORSCH, W. W. Multiple comparisons for analysing dichotomous response. *Biometrics*, Washington, v.47, n.1, p.45-52, 1990.

SHAFFER, J. P. A semi-bayesian study of Duncan's bayesian multiple comparison procedure. *J. Stat. Plann. Infer.*, Amsterdam, v.82, p.197-213, 1999.

STEEL, R. G. D.; TORRIE, J. H. *Principles and procedures of statistics.* 2.ed. New York: McGraw-Hill, 1980. 633p.

WILLIAMS, D. A. Test for differences between several small proportions. *J. R. Stat. Soc.*, London, v.37, n.3, p.421-434, 1988.

Received in 20.06.2011.

Approved after revised in 06.02.2012.