

# COMPARAÇÕES MÚLTIPLAS E TESTES SIMULTÂNEOS PARA PARÂMETROS BINOMIAIS DE $k$ POPULAÇÕES INDEPENDENTES

Nádia Giaretta BIASE<sup>1</sup>  
Daniel Furtado FERREIRA<sup>2</sup>

- **RESUMO:** Uma estratégia utilizada para comparar várias proporções binomiais é aplicar o teste F em uma análise de variância, seguida de testes de comparações múltiplas, se a hipótese nula global for rejeitada. No entanto, as pressuposições exigidas pelos testes nem sempre são atendidas, portanto, essa estratégia não é adequada. Entre os métodos alternativos para contornar o problema, destaca-se os testes assintóticos. Esse trabalho teve por objetivos propor um teste de comparações múltiplas assintóticas de proporções binomiais, baseado em formas quadráticas, e avaliar seu desempenho utilizando simulação Monte Carlo, e divulgar e avaliar o desempenho dos testes assintóticos  $G^2$  e  $X^2$  de Pearson para a hipótese de várias proporções binomiais. Foram simuladas amostras de Monte Carlo para diferentes configurações de  $k$  populações binomiais com parâmetros  $\pi_i$  e  $n_i$  para a  $i$ -ésima população,  $i = 1, 2, \dots, k$ . O teste  $X^2$  controlou o erro tipo I, em níveis iguais ou inferiores aos valores nominais, e apresentou desempenho superior ao do teste  $G^2$  que foi liberal, nas situações de pequenas amostras e maiores números de populações. Como o teste de comparações múltiplas assintóticas controlou o erro tipo I por experimento de forma conservativa e os resultados de poder foram bons, recomenda-se sua aplicação em situações reais.
- **PALAVRAS-CHAVE:** Simulação Monte Carlo; razão de verossimilhanças; formas quadráticas; procedimentos de comparações múltiplas.

## 1 Introdução

Em diferentes áreas do conhecimento, o pesquisador se depara com a necessidade de realizar inferências a respeito de várias proporções binomiais. Esse

---

<sup>1</sup>Faculdade de Ciências Integradas do Pontal – FACIP, Universidade Federal de Uberlândia – UFU, CEP: 38302-000, Ituiutaba, MG, Brasil. E-mail: [nadiabiase@yahoo.com.br](mailto:nadiabiase@yahoo.com.br)

<sup>2</sup>Departamento de Ciências Exatas, Universidade Federal de Lavras - UFLA, Caixa Postal 37, CEP: 37200-000, Lavras, MG, Brasil. E-mail: [danielff@dex.ufla.br](mailto:danielff@dex.ufla.br)

é o caso, por exemplo, dos ensaios de germinação (Raven et al., 2005) e dos ensaios de eberação cromossômica (Williams, 1988; Scott et al., 1983). Boa parte desses estudos é analisada de maneira clássica, realizando testes F em uma análise de variância, para comparar as várias proporções binomiais, acompanhada de um teste de comparações múltiplas, como por exemplo os testes de Tukey, Duncan, Student-Newman-Keuls (SNK), entre outros, para identificar onde ocorrem as diferenças, no caso da rejeição da hipótese nula global pelo teste F.

A adoção dessa estratégia não é adequada, em geral, por violar as pressuposições exigidas pelos testes aplicados, quais sejam, de normalidade e de homogeneidade das variâncias. O uso de modelos lineares generalizados é uma alternativa para tentar contornar o problema (McCulloch e Searle, 2001). Nessa alternativa, a família exponencial, que engloba o modelo binomial, é utilizada. Assim, a distribuição adequada dos dados é empregada, embora a distribuição das estatísticas dos testes sejam apenas assintóticas (Nelder e Wedderburn, 1972; McCulloch e Searle, 2001; Dobson e Barnett, 2008). Além disso, não existem procedimentos de comparações múltiplas fazendo uso desse tipo de análise relacionados à família exponencial.

Outra opção apropriada são os métodos de reamostragem com (*bootstrap*) e sem (permutação) reposição. No entanto, dos estudos envolvendo proporções binomiais destaca-se o trabalho de Biase e Ferreira (2006) no qual testes de comparações múltiplas foram propostos e avaliados os desempenhos desses testes. Não foi apresentado um teste para a hipótese nula global de igualdade das proporções binomiais. Foram obtidos desempenhos excelentes dos procedimentos de comparações múltiplas propostos. O problema é a inexistência em programas de análises estatísticas, sejam gratuitos ou pagos, de opções ou rotinas para a realização desse teste.

Soluções assintóticas para o teste global da igualdade de várias proporções binomiais são discutidas por Williams (1988) e Krishnamoorthy e Peng (2008). Entre os testes destacam-se o da razão de verossimilhanças  $G^2$  e o de  $X^2$  de Pearson. Agresti (2002) mostrou todas as deduções destes testes no contexto de tabelas de contingência ou da distribuição multinomial. Um outro teste para a igualdade de proporções binomiais a um valor de referência  $\pi_0$  pré-determinado é apresentado por Krishnamoorthy et al. (2004). Para identificar quais proporções binomiais causaram a rejeição da hipótese nula, esses autores apresentaram também intervalos de confiança simultâneos para proporções binomiais, cujos limites, inferior e superior, foram obtidos a partir da função beta.

No contexto de comparações de proporções binomiais, Piegorsch (1991) apresentou vários intervalos de confiança simultâneos. Primeiramente, ele considerou intervalos simultâneos para qualquer conjunto finito de contrastes utilizando a aproximação de Bonferroni aplicada ao intervalo de confiança de Wald, utilizando a distribuição normal acumulada. Em uma segunda etapa, o autor considerou um método implementado por Hochberg e Tamhane (1987) usando o intervalo de Wald juntamente com a distribuição da amplitude normal padronizada, para comparar todas as diferenças entre proporções por meio de intervalos de

confiança. Esses dois métodos apresentaram pobres desempenhos, tendo taxas de erro consideravelmente maior do que o valor nominal quando os tamanhos amostrais eram pequenos. Um melhor desempenho foi obtido utilizando um procedimento de estimação intervalar simultâneo utilizando a formulação de Jeffreys-Perks, motivada pela aproximação bayesiana de Beal (1987).

Agresti et al. (2008) também compararam pares de proporções binomiais independentes considerando intervalos de confiança simultâneos que utilizam a distribuição da amplitude estudentizada com uma estatística *escore*, utilizando o intervalo de Wald modificado e a distribuição normal. Esse intervalo apresentou desempenho semelhante ao método proposto por Piegorsch (1991), e portanto, parece ser um procedimento útil para obter intervalos de confiança simultâneos para vários parâmetros binomiais, embora Piegorsch (1991) tenha encontrado desempenho ruim para pequenas amostras.

Por essas razões o presente trabalho teve por objetivos divulgar e avaliar o desempenho dos testes assintóticos  $G^2$  e  $X^2$  de Pearson para a hipótese de várias proporções binomiais. Ademais, pretende-se propor um teste de comparações múltiplas assintóticas de proporções binomiais e avaliar o desempenho do teste proposto por meio de simulação Monte Carlo.

## 2 Metodologia

Considerando amostras aleatórias independentes de tamanhos  $n_1, n_2, \dots, n_k$  de  $k$  populações binomiais com probabilidade de sucesso do evento de interesse dados por  $\pi_1, \pi_2, \dots, \pi_k$ , cujas realizações amostrais são  $y_1, y_2, \dots, y_k$  pode se testar a hipótese  $H_0: \pi_1 = \pi_2 = \dots = \pi_k = \pi_0$ , sendo  $\pi_0$  não especificado, o que faz com que esse trabalho difira do trabalho desenvolvido por Krishnamoorthy et al. (2004). O primeiro teste considerado refere-se ao teste de razão de verossimilhanças e o segundo, a um teste baseado na distribuição assintótica de formas quadráticas. Procedimentos de comparações múltiplas foram derivados da teoria apresentada nesses testes.

### 2.1 Teste de razão de verossimilhanças

Para testar a hipótese da igualdade de  $k$  proporções binomiais independentes  $H_0: \pi_1 = \pi_2 = \dots = \pi_k = \pi_0$ , foi desenvolvido o teste da razão de verossimilhanças.

As funções de verossimilhanças, considerando os modelos probabilísticos binomiais independentes, sob  $H_1$  (modelo irrestrito  $\Omega$ ) e sob  $H_0$  (modelo restrito  $\Omega_0$ ) foram maximizadas e a estatística do teste resultante, na forma geral é dada por:

$$\Lambda = \frac{L_{\Omega_0}(\mathbf{y}|\pi_0)}{L_{\Omega}(\mathbf{y}|\boldsymbol{\pi})}, \quad (1)$$

em que  $\hat{\pi}_0$  é o estimador de máxima verossimilhança de  $\pi_0$  no modelo restrito  $\Omega_0$ , sob  $H_0$ , e  $\hat{\boldsymbol{\pi}} = [\hat{\pi}_1, \hat{\pi}_2, \dots, \hat{\pi}_k]^\top$  é o estimador do vetor de parâmetros no modelo irrestrito  $\Omega$ , sob  $H_1$ .

Foi utilizado o fato de que  $G^2 = -2\ln(\Lambda)$  segue assintoticamente uma distribuição qui-quadrado com  $\nu = k - 1$  graus de liberdade para realizar o teste (Mood et al., 1974).

A partir da função de verossimilhança do modelo irrestrito  $\Omega$  foi obtida a função suporte. Essa expressão foi derivada em relação a  $\pi_i$  e igualada a zero, resultando no seguinte estimador de máxima verossimilhança:

$$\hat{\pi}_i = \frac{Y_i}{n_i}, \quad i = 1, 2, \dots, k. \quad (2)$$

Para o modelo restrito  $\Omega_0$ , desenvolveu-se raciocínio análogo ao modelo irrestrito para obter o estimador de máxima verossimilhança de  $\pi_0$ , dado por:

$$\hat{\pi}_0 = \frac{\sum_{i=1}^k n_i \hat{\pi}_i}{n} = \frac{\sum_{i=1}^k Y_i}{n} \quad (3)$$

em que  $n = \sum_{i=1}^k n_i$ . O estimador  $\hat{\pi}_0$ , sob  $H_0$ , é a média ponderada dos estimadores  $\hat{\pi}_i$  de (2) de cada população, tendo como peso os tamanhos das amostras.

A estatística do teste, utilizando (1) é, portanto, dada por:

$$\Lambda = \prod_{i=1}^k \left( \frac{\hat{\pi}_0}{\hat{\pi}_i} \right)^{Y_i} \left[ \frac{1 - \hat{\pi}_0}{1 - \hat{\pi}_i} \right]^{n_i - Y_i} \quad (4)$$

Assim, tem-se que:

$$G^2 = 2 \left\{ \sum_{i=1}^k Y_i \ln \left( \frac{n Y_i}{n_i \sum_{j=1}^k Y_j} \right) + \sum_{i=1}^k (n_i - Y_i) \ln \left[ \frac{n(n_i - Y_i)}{n_i(n - \sum_{j=1}^k Y_j)} \right] \right\} \quad (5)$$

possui distribuição assintótica qui-quadrado com  $\nu = k - 1$  graus de liberdade, pois, no modelo completo há  $k$  parâmetros e no modelo reduzido, 1.

Se  $G^2 \geq \chi_{\alpha, k-1}^2$ , a hipótese de nulidade das proporções binomiais das  $k$  populações deve ser rejeitada. Esse teste de razão de verossimilhanças é conhecido como  $G^2$  na literatura (Williams, 1988) e foi utilizado para fins de comparação com a modificação proposta neste trabalho do teste  $X^2$  de Pearson, baseada-se em formas quadráticas.

## 2.2 Teste assintótico qui-quadrado

Na segunda abordagem para testar a hipótese nula de igualdade de  $k$  proporções binomiais independentes, foi considerado que o estimador de máxima verossimilhança  $\hat{\pi}$  de  $\pi$  é assintoticamente normal com média  $\pi$  e matriz de covariância  $\Sigma_{k \times k}$ . Assim, utilizando os resultados de Ferreira (2008), fundamentados na teoria multivariada e na distribuição de formas quadráticas, verifica-se que:

$$X^2 = (\hat{\pi} - \pi)^\top \Sigma^{-1} (\hat{\pi} - \pi) \quad (6)$$

possui distribuição qui-quadrado com  $k$  graus de liberdade.

Sob  $H_0$ , o vetor  $\pi$  será  $\pi_0 = [\pi_0, \pi_0, \dots, \pi_0]^\top$ , que, no entanto, é desconhecido. Assim, o estimador de máxima verossimilhança  $\hat{\pi}_0$  será utilizado para estimar a matriz  $\Sigma$  como alternativa. Portanto, a quantidade pivotal (6) utilizando tal estimador terá distribuição qui-quadrado assintótica com  $\nu = k - 1$  graus de liberdade. A matriz  $\Sigma$  foi determinada a partir do valor hipotético  $\pi_0$ , também estimado.

Considerando  $X_{ij}$  uma variável aleatória Bernoulli para o  $j$ -ésimo ensaio correspondente à  $i$ -ésima população binomial, sendo  $j = 1, 2, \dots, n_i$  e  $i = 1, 2, \dots, k$ , então  $Y_i = \sum_{j=1}^{n_i} X_{ij}$ . Para o desenvolvimento da expressão (6), foi necessário obter a matriz de covariâncias  $\Sigma$  que depende das expressões para as variâncias e covariâncias de  $Y_i$ . Assim,

$$\begin{aligned} E(Y_i) &= E \left[ \sum_{j=1}^{n_i} X_{ij} \right] = \sum_{j=1}^{n_i} E(X_{ij}) = \sum_{j=1}^{n_i} \pi_i \\ &= n_i \pi_i. \end{aligned}$$

e

$$\begin{aligned} V(Y_i) &= V \left( \sum_{j=1}^{n_i} X_{ij} \right) = \sum_{j=1}^{n_i} V(X_{ij}) = \sum_{j=1}^{n_i} \pi_i (1 - \pi_i) \\ &= n_i \pi_i (1 - \pi_i). \end{aligned}$$

Utilizando o estimador de máxima verossimilhança  $\hat{\pi}_i$  (2), então:

$$\begin{aligned} V(\hat{\pi}_i) &= V \left( \frac{Y_i}{n_i} \right) = \frac{1}{n_i^2} n_i \pi_i (1 - \pi_i) \\ &= \frac{\pi_i (1 - \pi_i)}{n_i}. \end{aligned}$$

Assim, a matriz de covariância  $\Sigma$  é dada por:

$$\Sigma = \begin{bmatrix} \frac{\pi_1(1-\pi_1)}{n_1} & 0 & \dots & 0 \\ 0 & \frac{\pi_2(1-\pi_2)}{n_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{\pi_k(1-\pi_k)}{n_k} \end{bmatrix},$$

uma vez que a covariância entre  $\hat{\pi}_i$  e  $\hat{\pi}_{i'}$ , para  $i \neq i'$  é nula, devido a independência.

Dessa forma, a quantidade pivotal (6) sob  $H_0$  pode ser simplificada em:

$$X^2 = [\hat{\pi}_1 - \pi_0, \hat{\pi}_2 - \pi_0, \dots, \hat{\pi}_k - \pi_0] \begin{bmatrix} \frac{n_1}{\pi_1(1-\pi_1)} & \dots & 0 \\ 0 & \ddots & 0 \\ \vdots & \vdots & \vdots \\ 0 & \dots & \frac{n_k}{\pi_k(1-\pi_k)} \end{bmatrix} \begin{bmatrix} \hat{\pi}_1 - \pi_0 \\ \hat{\pi}_2 - \pi_0 \\ \vdots \\ \hat{\pi}_k - \pi_0 \end{bmatrix}.$$

que expandida resulta em:

$$X^2 = \sum_{i=1}^k \frac{n_i(\hat{\pi}_i - \pi_0)^2}{\pi_i(1-\pi_i)} = \sum_{i=1}^k \frac{(Y_i - n_i\pi_0)^2}{n_i\pi_i(1-\pi_i)}.$$

Essa quantidade pivotal não pode ser usada em situações reais, pois depende de  $\pi_i$ , que é desconhecido. Desse modo, uma alternativa que pode ser considerada baseia-se em um procedimento análogo ao  $X^2$  de Pearson para populações multinomiais, também utilizado no contexto de binomiais por Williams (1988), exceto por uma pequena correção no denominador da expressão da estimativa. Nesse caso, consideram-se a hipótese nula  $\pi_1 = \pi_2 = \dots = \pi_k = \pi_0$ , e o estimador dado em (3). Assim, a estatística desse teste é dada por:

$$X^2 = \sum_{i=1}^k \frac{(Y_i - n_i\hat{\pi}_0)^2}{n_i\hat{\pi}_0(1-\hat{\pi}_0)}. \quad (7)$$

A estatística (7), sob  $H_0$ , segue uma distribuição qui-quadrado assintótica com  $\nu = k - 1$  graus de liberdade. Essa estatística é conhecida como  $X^2$  de Pearson, porém, neste trabalho ela foi obtida considerando argumentos teóricos totalmente diferentes do usual. Os desempenhos das estatísticas (5) e (7) serão avaliados na seção (3.1), contemplando inclusive situações de pequenas amostras ( $n < 20$ ), no qual os testes assintóticos podem apresentar problemas. Assim, por meio destas simulações pode-se verificar qual teste é melhor nessas circunstâncias, uma vez que na prática, não se sabe o limite que eles atingiram a equivalência.

### 2.3 Comparações múltiplas

Para realizar o teste de comparações múltiplas foi utilizado o resultado geral (6), devidamente modificado para contemplar o estimador  $\hat{\pi}_0$ . Assim, para testar

$$H_0 : \pi_i = \pi_{i'}, \quad i \neq i' = 1, 2, \dots, k \quad (8)$$

foi utilizada a seguinte estatística:

$$\chi_1^2 = \frac{(\hat{\pi}_i - \hat{\pi}_{i'})^2}{\text{var}(\hat{\pi}_i) + \text{var}(\hat{\pi}_{i'}) - 2\text{cov}(\hat{\pi}_i, \hat{\pi}_{i'})} \quad (9)$$

que possui distribuição assintótica qui-quadrado com  $\nu = k$  graus de liberdade, sendo que  $\text{var}(\hat{\pi}_i)$  e  $\text{var}(\hat{\pi}_{i'})$  são as variâncias dos estimadores  $\hat{\pi}_i$  e  $\hat{\pi}_{i'}$ , respectivamente e  $\text{cov}(\hat{\pi}_i, \hat{\pi}_{i'})$  é a covariância entre eles.

A hipótese (8) será rejeitada se  $\chi_1^2 > \chi_{\alpha, k}^2$ , em que  $\chi_{\alpha, k}^2$  é o quantil superior 100 $\alpha$ % da distribuição qui-quadrado com  $\nu = k$  graus de liberdade. As variâncias e a covariância de (9) são desconhecidas e precisam ser estimadas. As particularidades desses estimadores estão apresentadas na seção (3.2).

## 2.4 Simulações Monte Carlo

Para a validação dos desempenhos dos testes avaliados nesse trabalho foram realizadas simulações Monte Carlo. Foram geradas  $k$  populações binomiais com parâmetros  $\pi_i$  e  $n_i$  para a  $i$ -ésima população,  $i = 1, 2, \dots, k$ . Diferentes configurações desses parâmetros foram consideradas, sendo que, em cada configuração, 10.000 amostras de Monte Carlo foram simuladas. As taxas de erro tipo I e poder foram computadas conforme a situação. No caso do teste de razão de verossimilhanças e do teste de qui-quadrado baseado em formas quadráticas, as taxas de erro tipo I foram mensuradas apenas sob  $H_0$  completa e, no caso particular das comparações múltiplas, as taxas de erro tipo I por experimento foram consideradas, nas situações simuladas sob  $H_0$ , completa e parcial. O poder de todos os testes foi avaliado sob  $H_0$  parcial.

Foram consideradas duas situações distintas. Na primeira, foram consideradas populações com  $\pi_i$ 's idênticos ( $H_0$  completa) e fixados em 0,01, 0,1 e 0,5. Esses valores foram escolhidos em consonância com aqueles propostos por Biase e Ferreira (2006), para fins de comparação dos resultados via *bootstrap* com os do presente trabalho.

Na segunda, foram simuladas situações de  $H_0$  parcial, dadas por  $H_0 : \pi_1 = \pi_2 = \dots = \pi_i \neq \pi_{i+1} = \pi_{i+2} = \dots = \pi_k$ . Nesse caso, foram considerados dois grupos distintos e cuja diferença entre seus parâmetros foi fixada em  $\Delta = 0,01, 0,05, 0,1, 0,2, 0,3, 0,4, 0,5, 0,6, 0,7, 0,8$  e  $0,9$ . Optou-se por considerar apenas dois grupos em função da simplicidade e da existência de trabalhos na literatura com essa situação, que serviram de referência para comparação. O valor do parâmetro  $\pi_\ell$  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  $\ell = 1, 2, \dots, 5$  no caso de  $k = 10$ . Também foi considerada uma situação em que o valor  $\pi_\ell$  do primeiro grupo foi fixado em 0,3, 0,45 e 0,5 e os valores de  $\Delta = 0,01; 0,10$  e  $0,4$ .

Foram consideradas  $k = 2, 5$  e  $10$  populações e tamanhos de amostras  $n_i = 10, 30$  e  $100$ , para cada população,  $\forall i = 1, 2, \dots, k$ . Foram fixados os níveis nominais de significância  $\alpha = 1\%$  e  $5\%$ . Nas comparações dentro dos grupos, sob  $H_0$  parcial e nas comparações entre populações sob  $H_0$  completa, foram avaliadas

as taxas de erro tipo I por experimento, no caso de comparações múltiplas. Ainda sob  $H_0$  parcial, o poder do procedimento de comparações múltiplas foi avaliado nas 10.000 simulações, nos testes envolvendo médias dos diferentes grupos.

Também foram realizadas algumas simulações adicionais para avaliar as taxas de erro tipo I dos testes de razão de verossimilhanças e qui-quadrado com tamanhos amostrais ( $n$ ) variando de 5 em 5, de 10 até 100, considerando  $k = 5$ . E, para as comparações múltiplas, foram simuladas algumas situações para avaliar o poder e as taxas e erro tipo I por experimento sob  $H_0$  parcial, considerando  $n = 400, 800$  e  $1000$ ,  $k = 5$  e  $10$ ,  $\Delta = 0,01, 0,02, 0,03$  e  $0,04$  para  $\alpha = 5\%$ . Para esta situação, os valores de  $\pi$  do primeiro grupo foram fixados em  $0,05$  e os do segundo foram dados por  $0,05 + \Delta$ .

O teste binomial exato, considerando um nível nominal de significância de  $1\%$  para a hipótese  $H_0 : \alpha = 5\%$  vs  $H_1 : \alpha \neq 5\%$  e  $H_0 : \alpha = 1\%$  vs  $H_1 : \alpha \neq 1\%$  foi aplicado. Se a hipótese nula for rejeitada e a taxa de erro tipo I observada for considerada significativamente ( $p < 0,01$ ) inferior ao nível nominal, então, o teste será considerado conservativo; se a taxa de erro tipo I observada for considerada significativamente ( $p < 0,01$ ) superior ao nível nominal, o teste será considerado liberal e, finalmente, se a taxa de erro tipo I observada for considerada não significativamente ( $p < 0,01$ ) diferente do nível nominal, o teste será considerado exato. Se  $m$  representa o número de hipóteses nulas rejeitadas em  $N = 10.000$  simulações Monte Carlo para o nível nominal  $\alpha$ , então a estatística teste é obtida, usando a relação entre a distribuição binomial e a F, com probabilidade de sucesso  $\pi = \alpha$ , 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. Se  $F \leq F_{0,005}$  ou se  $F \geq F_{0,995}$ , a hipótese nula pode ser rejeitada, ao nível nominal de significância de  $1\%$ , em que  $F_{0,005}$  e  $F_{0,995}$  são os quantis da distribuição F, com  $\nu_1$  e  $\nu_2$  graus de liberdade.

### 3 Resultados e discussão

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

#### 3.1 Testes para a hipótese nula global

O desempenho dos testes de razão de verossimilhanças e qui-quadrado assintótico, baseado em formas quadráticas, foi avaliado e os resultados das taxas de erro tipo I e poder foram apresentados e discutidos na sequência.



### 3.1.1 Erro tipo I sob $H_0$ completa

Na Tabela 1 são apresentadas, em porcentagem, as taxas de erro tipo I sob  $H_0$  completa dos testes de razão de verossimilhanças e formas quadráticas, identificados por  $G^2$  e  $X^2$ , respectivamente. Esses testes foram avaliados em função do número de populações  $k$ , do tamanho da amostra  $n$  e dos valores dos parâmetros binomiais  $\pi$ , considerando o nível nominal de 5%.

Pode-se verificar em quase todas as situações que o teste  $X^2$  controlou o erro tipo I em nível inferior ou no máximo igual ao valor nominal de 5%, exceto para  $k = 2$ ,  $n = 100$  e  $\pi = 0,5$ , em que a taxa de erro tipo I superou significativamente o valor nominal, indicando comportamento liberal. Considerando  $\pi = 0,01$ , observou-se que para todas as configurações de  $n$  e  $k$  avaliadas, o teste  $X^2$  apresentou taxas significativamente ( $P < 0,01$ ) menores do que 5%. O teste  $X^2$  apresentou esse mesmo comportamento para  $n = 10$ ,  $k = 2$  e 5 e para  $\pi = 0,1$  e 0,5 e também para  $n = 10$  com  $k = 10$  e  $\pi = 0,1$ . Em todos esses casos mencionados o teste  $X^2$  foi considerado conservativo. Nas demais situações, o teste  $X^2$  apresentou taxas de erro tipo I iguais ao valor nominal.

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 ( $\pi$ ) para os testes de razão de verossimilhanças ( $G^2$ ) e formas quadráticas ( $X^2$ ) ao nível nominal de 5%

$k$	$n$	$\pi = 0,01$		$\pi = 0,1$		$\pi = 0,5$	
		$G^2$	$X^2$	$G^2$	$X^2$	$G^2$	$X^2$
2	10	0,01 <sup>+</sup>	0,00 <sup>+</sup>	4,66	0,88 <sup>+</sup>	5,21	3,99 <sup>+</sup>
2	30	0,45 <sup>+</sup>	0,03 <sup>+</sup>	7,22*	5,40	5,35	5,33
2	100	6,01*	1,34 <sup>+</sup>	5,02	4,99	5,68*	5,68*
5	10	0,04 <sup>+</sup>	0,04 <sup>+</sup>	5,12	2,55 <sup>+</sup>	6,49*	4,33 <sup>+</sup>
5	30	0,58 <sup>+</sup>	0,53 <sup>+</sup>	8,57*	4,94	5,63*	5,12
5	100	3,25 <sup>+</sup>	2,66 <sup>+</sup>	5,47	4,85	5,36	5,31
10	10	0,01 <sup>+</sup>	1,78 <sup>+</sup>	5,65*	3,81 <sup>+</sup>	7,29*	4,75
10	30	0,24 <sup>+</sup>	4,36 <sup>+</sup>	8,27*	4,81	6,17*	5,55
10	100	4,41 <sup>+</sup>	4,35 <sup>+</sup>	5,38	4,51	4,88	4,71

\* significativamente superior ao nível nominal de significância de 5% considerando uma confiança de 99%.

<sup>+</sup> significativamente inferior ao nível nominal de significância de 5% considerando uma confiança de 99%.

Para o teste  $G^2$ , houve um comportamento distinto em relação ao teste  $X^2$  para o controle do erro tipo I, pois ou o teste foi liberal ou exato, considerando  $\pi = 0,1$  ou 0,5. O teste foi liberal, ou seja, apresentou taxas significativamente superiores ao nível nominal de 5%, com  $k = 2$ , 5 e 10 para  $n = 30$ , considerando  $\pi = 0,1$  e também para  $n = 10$  e  $k = 10$ . Para  $\pi = 0,5$ , este teste foi liberal para  $k = 5$  e 10 e  $n = 10$  e 30. Com  $n = 100$ , o teste em questão foi considerado

liberal para  $k = 2$  e  $\pi = 0,5$ . Em todos os outros casos, o tamanho do teste  $G^2$  não foi significativamente diferente do valor nominal de 5%. Considerando  $\pi = 0,01$ , apenas para  $k = 2$  e  $n = 100$  o teste  $G^2$  foi liberal, para os demais valores de  $k$  e  $n$ , esse teste apresentou taxas de erro tipo I significativamente ( $P < 0,01$ ) inferiores a 5%.

De maneira geral, o teste  $X^2$  apresentou um melhor desempenho em relação ao controle do erro tipo I, pois em apenas um caso o teste foi considerado liberal, quando comparado ao teste  $G^2$ . Entretanto, apesar desse valor ser significativamente ( $P < 0,01$ ) diferente de 0,05, a diferença não foi expressiva. O teste  $G^2$  foi liberal principalmente nas situações de amostras pequenas e maior número de populações envolvidas. Em parte esses resultados são esperados em virtude da natureza assintótica da distribuição da estatística de um teste de razão de verossimilhanças. Além disso, o desempenho desses testes foi semelhante ao apontado por Williams (1988).

Procurando avaliar o comportamento dos testes para tamanhos amostrais variando entre 10 e 100, foram feitas simulações adicionais, considerando  $k = 5$ . Na Figura 1 estão apresentadas as taxas de erro tipo I observadas, sob  $H_0$  completa dos dois testes para  $k = 5$  com  $\pi = 0,1$  e  $0,5$  em função dos tamanhos amostrais, considerando  $\alpha = 5\%$ . De um modo geral, o teste  $G^2$  apresentou tamanho não significativamente ( $P > 0,01$ ) diferente do valor nominal ou então foi considerado liberal, ao contrário do teste  $X^2$ , que nas situações em que as taxas diferiram significativamente ( $P < 0,01$ ) do valor nominal, o teste foi considerado conservativo.

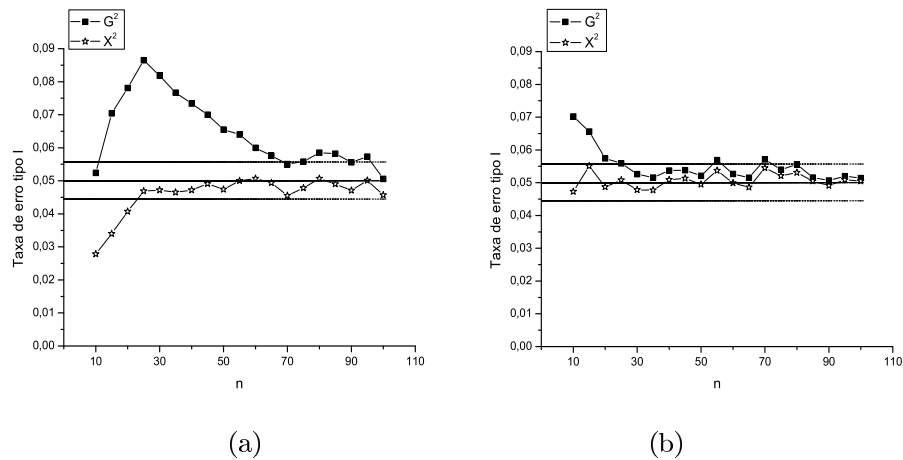


Figura 1 - Taxas de erro tipo I dos testes de razão de verossimilhanças ( $G^2$ ) e formas quadráticas ( $X^2$ ), em função dos tamanhos amostrais ( $n$ ) e valores dos parâmetros (a)  $\pi = 0,1$  e (b)  $\pi = 0,5$  para  $\alpha = 5\%$ , considerando a hipótese  $H_0$  completa.

Para pequenas amostras ( $n = 10$ ) na Figura 1 (a), a taxa de erro tipo I do teste  $G^2$  é não significativamente ( $P > 0,01$ ) diferente do valor nominal, e para tamanhos amostrais variando entre 15 e 60, em muitos casos, essas taxas além de serem significativamente diferentes de 5% são expressivamente elevadas. Com  $n \geq 65$ , as taxas de erro tipo I do teste tenderam a se igualar ao nível nominal de significância. Por outro lado, o teste  $X^2$  foi conservativo para amostras pequenas ( $n \leq 20$ ) Figura 1 (a), e para os demais tamanhos amostrais as taxas de erro tipo I passaram a ser significativamente iguais ao valor nominal de 5%.

Para  $\pi = 0,5$  na Figura 1 (b), o teste  $G^2$  foi considerado liberal para  $n \leq 25$ , 55 e 70. Nos demais casos, o teste apresentou tamanho não significativamente diferente do valor nominal. Independentemente dos tamanhos amostrais, as taxas de erro tipo I do teste  $X^2$  não diferiram do valor nominal.

Pode-se observar que o teste  $X^2$  apresentou desempenho um pouco melhor, pois na maioria das situações, as taxas de erro tipo I desse teste foram iguais ao nível nominal adotado e as do teste  $G^2$  foram superiores ao valor nominal em pequenas amostras, principalmente se  $\pi$  se afasta de 0,5.

### 3.1.2 Poder sob $H_0$ parcial

Na Tabela 2 são apresentados os valores de poder dos testes  $G^2$  e  $X^2$ , em função de  $k$ ,  $n$  e  $\Delta$ , considerando  $\alpha = 5\%$ , sob  $H_0$  parcial. É importante enfatizar que para medir os valores de poder sob a hipótese  $H_0$  parcial, considerou-se a formação de dois grupos. Os valores de  $\pi$  no primeiro grupo foram fixados em 0,01 e no segundo grupo esses valores, foram dados por  $0,01 + \Delta$ .

Para valores de  $\Delta = 0,01$ , os valores de poder de ambos os testes são inferiores ou relativamente próximos ao nível nominal de 5%, exceto para populações e amostras grandes ( $k = 5, 10$  e  $n = 100$ ). Se os valores de  $\Delta$  são pequenos ou moderados ( $\Delta = 0,1$  e  $0,3$ ), observa-se que o poder dos testes, em algumas situações é pequeno, considerando tamanhos amostrais pequenos ou intermediários. Isso foi constatado para  $\Delta = 0,1$  com  $n = 10$  e 30 com  $k = 5$  e  $n = 10$  com  $k = 10$  para ambos os testes. Para grandes amostras ( $n = 100$ ), o desempenho dos testes se igualou e seus valores de poder aproximaram de 100%. Para  $\Delta = 0,3$  o poder foi pequeno apenas para o teste  $X^2$  com  $k = 5$  e  $n = 10$ . Nos demais casos, o poder de ambos os testes atingiu 100% rapidamente.

De maneira geral, verifica-se que o teste  $G^2$  apresentou valores de poder quase sempre superiores ao do teste  $X^2$ , e que com o aumento do tamanho das amostras, o poder dos testes tende a aumentar expressivamente, como já era esperado. Um outro aspecto interessante a ser destacado, refere-se ao desempenho de ambos os testes quando o número de populações  $k$  eleva-se de 5 para 10. Nesses casos, observa-se que os valores de poder tendem a ser maiores, se for fixado o valor de  $\Delta$ ,  $n$  e o teste, porém, esse aumento foi sutilmente menor, quando os tamanhos das amostras e valores de  $\Delta$  são grandes ( $n \geq 30$  e  $\Delta \geq 0,3$ ).

Para diferenças grandes ou muito grandes ( $\Delta \geq 0,4$ ), verifica-se que o poder de ambos os testes aproximam-se de 100%, mesmo para pequenas amostras ( $n = 10$ ).

Tabela 2 - Poder (%) sob  $H_0$  parcial para diferentes números de populações ( $k$ ), de tamanhos de amostras ( $n$ ) e de diferenças entre as proporções binomiais ( $\Delta$ ) para os testes de razão verossimilhanças ( $G^2$ ) e formas quadráticas ( $X^2$ ) ao nível nominal de 5%

$k$	$n$	$\Delta = 0,01$		$\Delta = 0,1$		$\Delta = 0,3$		$\Delta = 0,4$	
		$G^2$	$X^2$	$G^2$	$X^2$	$G^2$	$X^2$	$G^2$	$X^2$
5	10	0,12	0,11	9,18	6,73	77,22	64,33	94,39	88,46
5	30	1,90	1,54	64,61	50,36	99,95	99,93	100,00	100,00
5	100	9,05	6,86	99,54	99,48	100,00	100,00	100,00	100,00
10	10	0,05	3,20	13,18	16,82	95,29	86,31	99,82	98,62
10	30	1,22	6,24	86,58	74,03	100,00	100,00	100,00	100,00
10	100	15,01	10,24	100,00	100,00	100,00	100,00	100,00	100,00
$k$	$n$	$\Delta = 0,5$		$\Delta = 0,7$		$\Delta = 0,8$		$\Delta = 0,9$	
		$G^2$	$X^2$	$G^2$	$X^2$	$G^2$	$X^2$	$G^2$	$X^2$
5	10	98,92	97,52	100,00	100,00	100,00	100,00	100,00	100,00
5	30	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00
5	100	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00
10	10	99,99	99,98	100,00	100,00	100,00	100,00	100,00	100,00
10	30	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00
10	100	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00

Assim, se o pesquisador tem interesse em averiguar pequenas diferenças ( $\Delta < 0,01$ ) entre populações de grupos diferentes, é aconselhável estabelecer tamanhos de amostras maiores que 100, principalmente se o número de populações for menor.

O desempenho do poder dos testes estudados também foi avaliado em algumas situações em que os valores de  $\pi$  do primeiro e segundo grupo estivessem próximos de 0,5, pois nesses casos, os testes de hipóteses tradicionais que envolvem proporções binomiais apresentam melhores propriedades. Assim foram realizadas algumas simulações em que os valores de  $\pi$  do primeiro grupo, denominado de  $\pi^{(1)}$ , foram fixados em 0,30, 0,45, e 0,50 e os do segundo grupo foram dados pela combinação dos valores de  $\pi^{(1)}$  e  $\Delta$ , ou seja,  $\pi^{(2)} = \pi^{(1)} + \Delta$ . Nessas situações específicas os valores de  $\Delta$  foram iguais a 0,01, 0,1 e 0,4.

Na Tabela 3 são apresentados os valores de poder dos testes  $G^2$  e  $X^2$  em função de  $k$ ,  $\pi^{(1)}$ ,  $n$  e  $\Delta$ , considerando o nível nominal de 5%. Pode-se observar, que para  $\Delta = 0,01$  os valores de poder de ambos os testes são relativamente pequenos, independente dos tamanhos amostrais  $n$  e do número de populações  $k$ . Para  $\Delta = 0,1$ , verifica-se que os valores de poder dos dois testes também foram pequenos, mas com o aumento do tamanho das amostras o poder dos testes teve um crescimento considerável, principalmente se  $n$  aumenta de 30 para 100.

O desempenho dos testes foi consideravelmente melhor para  $\Delta = 0,4$ , apresentando valores de poder pequenos apenas para  $k = 5$  e  $n = 10$ . Para amostras intermediárias e grandes ( $n \geq 30$ ), os valores de poder dos testes tenderam a se igualar e se aproximaram de 100%.

Observou-se que se a diferença entre os valores de  $\pi$  é pequena ( $\Delta = 0,01$ ), o poder dos testes é pequeno, e à medida que essa diferença torna-se maior ( $\Delta = 0,1$  e  $0,4$ ), o poder dos testes tem um aumento avultado, principalmente para grandes amostras ( $n \geq 30$ ). Isso pode ser verificado na Tabela 3, fixando um valor de  $k$ ,  $\pi^{(1)}$ ,  $n$  e o teste. Finalmente, pode-se observar um grande aumento do poder dos testes com o aumento de  $k = 5$  para  $k = 10$ . Esse comportamento também foi semelhante ao observado para valores de  $\pi$  afastado de  $0,5$  (Tabela 2).

Se as situações de  $\pi^{(1)} = 0,01$  e  $\pi^{(1)} = 0,30, 0,45$  e  $0,50$  (Tabelas 2 e 3) forem comparadas, pode-se observar que o poder aumenta com o aumento da diferença paramétrica  $\Delta$  entre as populações, fixados  $k$  e  $n$ , como já é preconizado pela teoria. No entanto, ao contrário do que se esperava, quando  $\pi^{(1)}$  se aproxima de  $0,5$  (Tabela 3) o aumento do valor de poder com o aumento de  $\Delta$  é menor do que esse mesmo aumento com  $\pi^{(1)} = 0,01$  (Tabela 2). Simulações adicionais mostraram que isso ocorre também para valores de  $\pi^{(1)}$  variando de  $0,01$  a  $0,35$  (Resultados não apresentados). A partir desse ponto ( $\pi^{(1)} = 0,35$ ) para um dado valor fixo de  $\Delta$  o poder é maior na medida que  $\pi^{(1)}$  aumenta e se aproxima de  $0,5$ .

Tabela 3 - Poder (%) sob  $H_0$  parcial para diferentes números de populações ( $k$ ), de valores do parâmetro  $\pi$  no primeiro grupo ( $\pi^{(1)}$ ), de tamanhos de amostras ( $n$ ) e diferenças entre os parâmetros  $\pi$  de cada grupo ( $\Delta$ ) para os testes de razão verossimilhanças ( $G^2$ ) e formas quadráticas ( $X^2$ ) ao nível nominal de 5%

$k$	$\pi^{(1)}$	$n$	$\Delta = 0,01$		$\Delta = 0,1$		$\Delta = 0,4$	
			$G^2$	$X^2$	$G^2$	$X^2$	$G^2$	$X^2$
5	0,30	10	9,00	5,05	10,73	6,82	63,94	58,24
5	0,30	30	5,66	5,07	15,23	14,45	98,81	98,71
5	0,30	100	5,37	5,24	42,32	42,17	100,00	100,00
10	0,30	10	9,14	4,93	12,50	7,64	85,39	81,06
10	0,30	30	6,08	4,93	19,29	17,47	100,00	100,00
10	0,30	100	6,12	5,79	61,95	61,35	100,00	100,00
5	0,45	10	7,38	4,86	9,38	6,67	70,96	61,98
5	0,45	30	5,19	4,67	14,19	13,44	99,37	99,30
5	0,45	100	5,33	5,20	38,93	38,71	100,00	100,00
10	0,45	10	6,75	4,40	11,16	7,16	90,57	86,48
10	0,45	30	5,74	5,10	18,70	17,15	100,00	100,00
10	0,45	100	5,24	5,03	57,39	56,73	100,00	100,00
5	0,50	10	6,74	4,47	9,92	6,89	76,22	66,11
5	0,50	30	5,71	5,09	14,22	13,44	99,78	99,75
5	0,50	100	5,44	5,26	39,36	38,91	100,00	100,00
10	0,50	10	6,83	4,35	11,21	7,64	94,33	90,37
10	0,50	30	5,57	4,95	18,22	16,75	100,00	100,00
10	0,50	100	5,15	5,01	56,95	56,17	100,00	100,00

Para se ter uma percepção mais nítida do que ocorreu com os valores de poder quando os valores de  $\pi^{(1)}$  se aproximam de 0,5, o poder de ambos os testes são apresentados na Figura 2 em função de  $\Delta$  com  $k = 5$  e  $n = 10$ , considerando  $\alpha = 5\%$ . Pode se observar que o poder de ambos os testes, considerando  $\Delta = 0,01$  e  $\pi^{(1)} = 0,01$ , é inferior e relativamente próximo de zero em relação aos valores de  $\pi^{(1)}$  próximos de 0,5 ( $\pi^{(1)} = 0,30, 0,45$  e  $0,50$ ).

Para  $\Delta = 0,1$ , os valores de poder de ambos os testes são pequenos e tendem a se igualar para diferentes valores de  $\pi^{(1)}$ . Por fim, considerando  $\Delta = 0,4$ , verifica-se que houve uma inversão dos valores de poder quanto aos valores de  $\pi^{(1)}$ , ou seja, para uma diferença maior de  $\Delta$ , os valores de poder dos testes para  $\pi^{(1)} = 0,01$  foram relativamente superiores se comparado com os demais valores de  $\pi^{(1)}$  avaliados. Quanto mais afastados do 0,5 forem os valores de  $\pi^{(1)}$  maiores são os valores de poder.

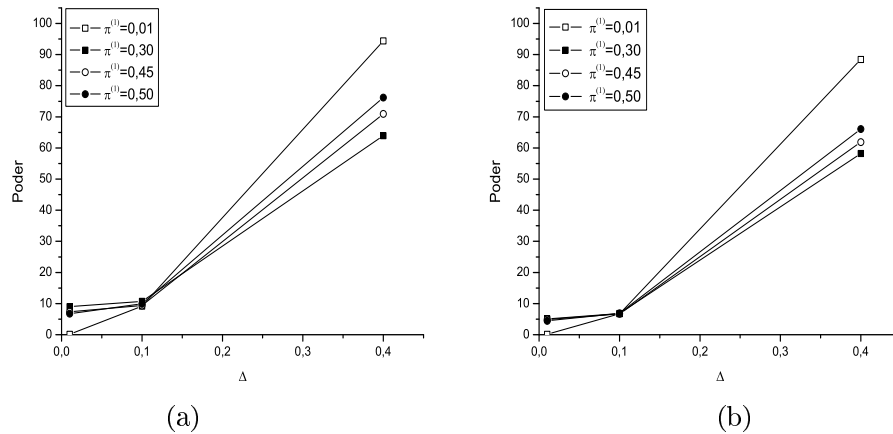


Figura 2 - Poder sob  $H_0$  parcial, em função da diferença  $\Delta$ , com  $k = 5$ ,  $n = 10$ , considerando  $\alpha = 5\%$  dos testes (a)  $G^2$  e (b)  $X^2$ .

A superioridade do teste  $G^2$  em relação ao teste  $X^2$  considerando o poder tem de ser vista com certas ressalvas. Devido ao fato de que as taxas de erro tipo I do teste  $G^2$  são superiores aos níveis nominais em pequenas amostras, e valores de  $\pi$  afastados de 0,5. Essas situações são exatamente aquelas em que sob  $H_1$  o teste  $G^2$  apresentou maior poder. Assim, numa situação real, em que o pesquisador não sabe se está sob  $H_0$  ou sob  $H_1$ , o risco de rejeitar  $H_0$ , sendo esta verdadeira é maior que o nível nominal  $\alpha$ , pois o teste  $G^2$  é liberal. Por essa razão as vantagens em relação ao poder desse teste devem ser ponderadas pela desvantagem maior de se cometer o erro tipo I.

### 3.2 Comparações múltiplas

A estatística (9) para testar a hipótese (8), utilizando o estimador da variância, é simplificada por:

$$\chi_1^2 = \frac{(\hat{\pi}_i - \hat{\pi}_{i'})^2}{\frac{\hat{\pi}_i(1-\hat{\pi}_i)}{n_i} + \frac{\hat{\pi}_{i'}(1-\hat{\pi}_{i'})}{n_{i'}}}, \quad (10)$$

uma vez que  $\text{vâr}(\hat{\pi}_i) = \hat{\pi}_i(1-\hat{\pi}_i)/n_i$ ,  $\text{vâr}(\hat{\pi}_{i'}) = \hat{\pi}_{i'}(1-\hat{\pi}_{i'})/n_{i'}$  e a  $\text{côv}(\hat{\pi}_i, \hat{\pi}_{i'}) = 0$ . Sob a hipótese nula,  $\chi_1^2$  possui distribuição assintótica qui-quadrado com  $\nu = k - 1$  graus de liberdade. Se  $\chi_1^2 \geq \chi_{\alpha, k}^2$  a hipótese (8) deve ser rejeitada.

No entanto, essa expressão possui muitas limitações práticas, principalmente se  $n_i$ 's for pequeno e se os valores de  $\pi_i$ 's se aproximam de 0 ou de 1. A razão disso é que pode ocorrer denominadores nulos na expressão (10) em decorrência de estimativas nulas das variâncias. Assim para contornar essa limitação foi proposto o estimador comum (3) dos  $\pi_i$ 's para ser utilizado no estimador das variâncias dos  $\hat{\pi}_i$ 's. Assim, a estatística do teste foi dada por:

$$\chi_2^2 = \frac{(\hat{\pi}_i - \hat{\pi}_{i'})^2}{\frac{\hat{\pi}_0(1-\hat{\pi}_0)}{n_i} + \frac{\hat{\pi}_0(1-\hat{\pi}_0)}{n_{i'}}}, \quad (11)$$

que sob  $H_0$ , segue uma distribuição de qui-quadrado com  $\nu = k - 1$  graus de liberdade.

Nesse caso, as estimativas de  $\pi_0$  podem ser nulas, ou iguais a unidade em algumas situações, o que levaria o denominador de (11) a ser nulo. Nesse caso, haveria uma impossibilidade matemática, pois não existe divisão por zero. Como  $\hat{\pi}_0$  seria igual a 0 se todas as realizações das  $k$  populações, fossem iguais a 0, é natural considerar  $\chi_1^2$  como nulo. Da mesma forma,  $\chi_2^2$  seria nulo se  $\hat{\pi}_0$  fosse igual a 1, que ocorre quando todas as realizações  $y_i$  forem iguais a  $n_i$ . Em ambos os casos todos os estimadores seriam iguais, o que é uma forte evidência a favor de  $H_0$ .

Inicialmente a avaliação do desempenho do teste de comparações múltiplas foi realizada considerando o erro tipo I por experimento, mensurado sob a hipótese  $H_0$  completa e sob a hipótese  $H_0$  parcial, nas comparações intragrupos. Posteriormente, o poder do TCM foi avaliado nas comparações intergrupos sob a hipótese  $H_0$  parcial.

#### 3.2.1 Erro tipo I sob $H_0$ completa

Na Tabela 4 são apresentadas, em porcentagem, as taxas de erro tipo I por experimento sob  $H_0$  completa para o teste de comparações múltiplas, denotado por TCM, em função de  $k$ ,  $n$  e  $\pi$ , considerando o nível nominal de 5%. O que se observa, de maneira geral, é que houve controle do erro tipo I por experimento, que ocorreu de forma conservativa, independentemente do número de populações, dos tamanhos amostrais e dos valores de  $\pi$ .

Biase e Ferreira (2006) avaliaram as taxas de erro tipo I por experimento em situações similares as consideradas neste trabalho sob  $H_0$  completa, para dois

Tabela 4 - 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 de  $\pi$  para o teste de comparações múltiplas (TCM) ao nível nominal de 5%

$k$	$n$	$\pi = 0,01$	$\pi = 0,1$	$\pi = 0,5$
5	10	0,00 <sup>+</sup>	0,31 <sup>+</sup>	2,14 <sup>+</sup>
5	30	0,03 <sup>+</sup>	1,30 <sup>+</sup>	2,10 <sup>+</sup>
5	100	0,32 <sup>+</sup>	1,45 <sup>+</sup>	1,86 <sup>+</sup>
10	10	0,00 <sup>+</sup>	0,06 <sup>+</sup>	0,04 <sup>+</sup>
10	30	0,00 <sup>+</sup>	0,10 <sup>+</sup>	0,20 <sup>+</sup>
10	100	0,13 <sup>+</sup>	0,10 <sup>+</sup>	0,09 <sup>+</sup>

\* significativamente superior ao nível nominal de significância de 5% considerando uma confiança de 99%.

<sup>+</sup> significativamente inferior ao nível nominal de significância de 5% considerando uma confiança de 99%.

testes de *bootstrap*, que se diferenciaram pelo estimador das proporções utilizado, sendo um deles o estimador de máxima verossimilhança e o outro o estimador add-4 de Pan (Pan, 2002). Considerando as mesmas configurações para  $k$ ,  $n$  e  $\pi = 0,1$  e  $0,5$  com  $\alpha = 5\%$ , esses dois testes de *bootstrap* controlaram o erro tipo I por experimento em todas as situações avaliadas e apresentaram muitas taxas significativamente inferiores ao valor nominal para pequenas amostras ( $n = 10$  ou  $30$ ) e valores de  $\pi$  afastados de  $0,5$ . Para grandes amostras ( $n = 100$ ), os dois testes foram considerados exatos.

O TCM apresentou algumas diferenças em relação ao comportamento dos testes de *bootstrap* sendo considerado conservativo em todas as situações avaliadas. Isso não ocorreu com os testes de *bootstrap*. No entanto, é importante salientar que os testes de *bootstrap* exigem maior esforço computacional se comparado com o TCM avaliado nesse trabalho. Esse fato deve ser levado em consideração pelo pesquisador quando seu interesse for realizar comparações múltiplas envolvendo proporções binomiais. Ademais rotinas para a aplicação destes testes de *bootstrap* ainda não são encontradas nos programas de análise estatísticas. Isso é um grande limitador para muitos pesquisadores, que em geral possuem dificuldades de implementar rotinas como as exigidas para a aplicação dos testes de *bootstrap*.

Piegorsch (1991) avaliando o desempenho dos intervalos simultâneos, utilizando o método aproximado de Bonferroni e o de Hochberg e Tamhane para comparar proporções binomiais, verificou também que para pequenas amostras esses métodos apresentaram taxas de erro tipo superiores ao valor nominal. Resultados semelhantes foram encontrados por Agresti et al. (2008) na comparação de pares de proporções binomiais utilizando o método da distribuição da amplitude estudentizada com a estatística *escore*.



### 3.2.2 Erro tipo I sob $H_0$ parcial

Na Tabela 5, as taxas de erro tipo I por experimento sob  $H_0$  parcial para o TCM são apresentadas em função de  $k$ ,  $n$  e  $\Delta$  para  $\alpha = 5\%$ . Verificou-se que em todos os casos o TCM controlou o erro tipo I por experimento e foi classificado como conservativo em todas as situações. O TCM apresentou excesso de conservadorismo para valores de  $\pi$  de um ou ambos os grupos afastados de 0,5 ( $\Delta \leq 0,05$  ou  $\Delta \geq 0,6$ ), lembrando que, para avaliar as taxas de erro tipo I sob  $H_0$  parcial, estabeleceu-se que os valores de  $\pi$  no primeiro grupo fossem iguais a 0,01 e de  $0,01 + \Delta$  no segundo grupo. Este artifício foi utilizado para evitar que os valores de  $\pi$  de ambos os grupos extrapolassem o espaço paramétrico dos parâmetros binomiais  $\pi$ , que variam entre 0 e 1.

Tabela 5 - 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  $\pi$  de cada grupo ( $\Delta$ ), para o teste de comparações múltiplas (TCM) ao nível nominal de 5%

$k$	$n$	$\Delta$								
		0,01	0,05	0,1	0,3	0,4	0,5	0,6	0,7	0,9
5	10	0,00 <sup>+</sup>	0,18 <sup>+</sup>	0,80 <sup>+</sup>	2,66 <sup>+</sup>	1,91 <sup>+</sup>	1,35 <sup>+</sup>	0,57 <sup>+</sup>	0,27 <sup>+</sup>	0,00 <sup>+</sup>
5	30	0,14 <sup>+</sup>	1,93 <sup>+</sup>	2,81 <sup>+</sup>	2,65 <sup>+</sup>	1,82 <sup>+</sup>	1,39 <sup>+</sup>	0,64 <sup>+</sup>	0,26 <sup>+</sup>	0,00 <sup>+</sup>
5	100	0,71 <sup>+</sup>	2,30 <sup>+</sup>	3,13 <sup>+</sup>	2,61 <sup>+</sup>	1,81 <sup>+</sup>	1,19 <sup>+</sup>	0,69 <sup>+</sup>	0,22 <sup>+</sup>	0,00 <sup>+</sup>
10	10	0,00 <sup>+</sup>	0,08 <sup>+</sup>	0,49 <sup>+</sup>	0,66 <sup>+</sup>	0,38 <sup>+</sup>	0,26 <sup>+</sup>	0,08 <sup>+</sup>	0,02 <sup>+</sup>	0,00 <sup>+</sup>
10	30	0,09 <sup>+</sup>	0,78 <sup>+</sup>	1,14 <sup>+</sup>	0,90 <sup>+</sup>	0,58 <sup>+</sup>	0,38 <sup>+</sup>	0,10 <sup>+</sup>	0,02 <sup>+</sup>	0,00 <sup>+</sup>
10	100	0,16 <sup>+</sup>	1,07 <sup>+</sup>	1,24 <sup>+</sup>	0,99 <sup>+</sup>	0,61 <sup>+</sup>	0,24 <sup>+</sup>	0,07 <sup>+</sup>	0,01 <sup>+</sup>	0,00 <sup>+</sup>

<sup>+</sup> significativamente inferior ao nível nominal de significância de 5% considerando uma confiança de 99%.

Comparando as taxas de erro tipo I para  $k = 5$  e  $k = 10$ , observa-se para um valor fixo de  $\Delta$ , que houve redução das taxas quando o número de populações aumentou de 5 para 10, independentemente dos tamanhos amostrais. Esse comportamento foi semelhante ao observado para as taxas de erro tipo I por experimento sob  $H_0$  completa.

Os resultados das taxas de erro tipo I, por experimento, do TCM para  $\alpha = 5\%$  são coerentes com os resultados obtidos para os testes de *bootstrap* avaliados por Biase e Ferreira (2006), que também apresentaram na grande maioria dos casos, taxas de erro tipo I por experimento não significativamente ( $P > 0,01$ ) diferentes ou significativamente ( $P < 0,01$ ) inferiores ao valor nominal adotado. Portanto, tanto os testes de *bootstrap* como o TCM controlou o erro tipo I sob a hipótese  $H_0$  parcial e apresentou comportamento bastante similar para todos os tamanhos de amostras  $n$ , número de populações  $k$  e valores de  $\Delta$ .

### 3.2.3 Poder sob $H_0$ parcial

Na Tabela 6 são apresentados os valores de poder, em porcentagem, para o TCM, em função de  $k$ ,  $n$  e  $\Delta$  para  $\alpha = 5\%$ . Para valores muito pequenos de  $\Delta$

( $\Delta = 0,01$  e  $0,05$ ), observa-se que os valores de poder são extremamente pequenos, independente do número de populações e do tamanho das amostras. Para valores pequenos e moderados de  $\Delta$  ( $\Delta = 0,1$  e  $0,3$ ), os valores de poder do TCM aumentam consideravelmente com o aumento de  $n$ , principalmente de 30 para 100, com exceção para  $k = 10$  e  $n = 100$  considerando  $\Delta = 0,1$ , em que o poder do TCM não apresentou um grande crescimento.

Tabela 6 - Poder (%) sob  $H_0$  parcial para diferentes números de populações ( $k$ ), tamanhos de amostras ( $n$ ) e diferenças entre os parâmetros  $\pi$  de cada grupo ( $\Delta$ ), para o teste de comparações múltiplas (TCM) ao nível nominal de 5%

$k$	$n$	$\Delta$								
		0,01	0,05	0,1	0,3	0,4	0,5	0,6	0,7	0,9
5	10	0,00	0,09	0,48	8,50	18,59	33,77	51,10	69,69	98,62
5	30	0,07	1,53	7,46	67,72	89,83	98,09	99,88	100,00	100,00
5	100	0,53	14,83	56,63	99,99	100,00	100,00	100,00	100,00	100,00
10	10	0,00	0,02	0,10	0,40	0,94	1,91	4,91	11,79	35,03
10	30	0,02	0,16	0,57	13,89	35,34	65,37	89,55	98,37	100,00
10	100	0,03	1,25	10,96	97,51	99,97	100,00	100,00	100,00	100,00

Para  $\Delta = 0,4$ , verificou-se que os valores de poder do TCM foram pequenos para pequenas amostras ( $n = 10$ ) e para  $k = 10$  e  $n = 30$ . Para grandes amostras os valores de poder do TCM aproximaram de 100%. Considerando  $\Delta = 0,5$  e  $0,6$ , o poder do TCM foi pequeno apenas para pequenas amostras ( $n = 10$ ), tanto para  $k = 5$  como para  $k = 10$ . Para amostras intermediárias ou grandes ( $n \geq 30$ ), os valores de poder do TCM são altos e tenderam a se aproximar de 100%. Se as diferenças ( $\Delta$ ) são muito grandes ( $\Delta \geq 0,7$ ), os valores de poder do TCM são altos e tendem a aproximar de 100%, exceto para  $k = 10$  e  $n = 10$ .

De maneira geral, verifica-se que para todos os valores de  $\Delta$ , houve uma redução expressiva dos valores de poder com o aumento do número de populações  $k$ , e que o aumento de  $n$  propicia incrementos consideráveis no poder do TCM.

Biase e Ferreira (2006) avaliando o poder dos testes de *bootstrap* sob a hipótese  $H_0$  parcial, concluiu também que os valores de poder dos testes de *bootstrap* sofrem uma grande redução com o aumento do número de populações  $k$  e que com o aumento dos tamanhos amostrais  $n$  o poder dos testes tem um crescimento expressivo, tanto para o teste MV como para o teste Pan. O mesmo efeito do poder do TCM avaliado neste trabalho foi verificado com o aumento de  $n$  e  $k$ . O mesmo comportamento é observado nas comparações múltiplas de médias normais para o teste Tukey. Nesse caso, o teste se torna muito conservativo com grandes valores de  $k$  sob  $H_0$  e pouco poderoso, na mesma situação, sob  $H_1$  (Borges e Ferreira, 2003; Santos et al., 2001)

Na Tabela 7 são apresentados os valores de poder do TCM nas situações em que os valores de  $\pi$  dos grupos se aproximavam de 0,5, em função de  $k$ ,  $\pi^{(1)}$ ,  $n$  e  $\Delta$  para  $\alpha = 5\%$ . Pode-se observar para  $\Delta = 0,01$  que o poder do TCM foi expressivamente pequeno, e que em todas as situações considerando  $k = 10$ , o poder foi inferior ao

valor nominal de 5%. Para  $\Delta = 0,1$ , os valores de poder foram pequenos mesmo para grandes amostras ( $n = 100$ ). Para uma diferença entre os grupos um pouco maior ( $\Delta = 0,4$ ), os valores de poder foram pequenos com  $n \leq 30$ , e para grandes amostras ( $n = 100$ ), os valores de poder do TCM aproximaram de 100%.

De modo geral, verifica-se que se os valores de  $\pi$  dos dois grupos ( $\pi^{(1)}$  e  $\pi^{(2)}$ ) são muito próximos ( $\Delta = 0,01$  e  $0,1$ ), o poder do teste é pequeno, independente do número de populações  $k$  e dos tamanhos amostrais  $n$ . Para valores de ( $\pi^{(1)}$  e  $\pi^{(2)}$ ) mais bem diferenciados ( $\Delta = 0,4$ ), os valores de poder tendem a se aproximar de 100%, com o aumento do tamanho das amostras. E novamente, verifica-se que o poder do TCM sofre uma redução com o aumento de  $k = 5$  para  $k = 10$ . Esses resultados também estão de acordo com os apresentados por Biase e Ferreira (2006), considerando os testes de *bootstrap*.

Tabela 7 - Poder (%) sob  $H_0$  parcial para diferentes números de populações ( $k$ ), de valores do parâmetro  $\pi$  no primeiro grupo ( $\pi^{(1)}$ ), de tamanhos de amostras ( $n$ ) e diferenças entre os parâmetros  $\pi$  de cada grupo ( $\Delta$ ), para o teste de comparações múltiplas (TCM) ao nível nominal de 5%

$k$	$\pi^{(1)}$	$n$	$\Delta = 0,01$	$\Delta = 0,1$	$\Delta = 0,4$
5	0,30	10	0,13	0,43	10,88
5	0,30	30	0,19	1,04	56,05
5	0,30	100	0,23	5,72	99,74
10	0,30	10	0,00	0,00	0,11
10	0,30	30	0,01	0,02	16,41
10	0,30	100	0,00	0,42	94,58
5	0,45	10	0,24	0,52	9,66
5	0,45	30	0,30	1,29	56,52
5	0,45	100	0,25	5,14	99,85
10	0,45	10	0,00	0,00	0,42
10	0,45	30	0,01	0,05	15,99
10	0,45	100	0,00	0,27	97,80
5	0,50	10	0,25	0,47	8,53
5	0,50	30	0,31	1,13	58,72
5	0,50	100	0,25	4,87	99,97
10	0,50	10	0,00	0,00	0,43
10	0,50	30	0,01	0,04	21,30
10	0,50	100	0,00	0,37	98,89

É importante comentar que os tamanhos amostrais utilizados na experimentação agrícola são compatíveis com os utilizados no presente trabalho. No entanto, esses tamanhos amostrais são insuficientes para ter alto poder em detectar diferenças pequenas ( $\Delta < 0,10$ ). Por exemplo, nas pesquisas eleitorais são amostrados  $n = 1068$  eleitores para se ter um erro de 3 pontos percentuais com 95% de confiança. Nesse caso, pode-se afirmar que um candidato possui maior proporção de eleitores

votantes que outro, quando a diferença nas estimativas dessas proporções de ambos os candidatos for superior a 6 pontos percentuais. Caso essa diferença seja menor que esse valor, o resultado é declarado como “empate técnico”.

Esse exemplo ilustra uma situação em que uma amostra de tamanho  $n = 1068$  é necessária para detectar uma diferença de  $\Delta = 0,06$ . Esse valor é muito superior aos tamanhos amostrais da maioria das situações agropecuárias. Sugere-se que tamanhos amostrais bem maiores dos que são praticados normalmente sejam utilizados nas pesquisas.

Em decorrência desse fato, foram realizadas algumas simulações extras para avaliar o TCM em relação as taxas de erro tipo I por experimento e poder sob  $H_0$  parcial, considerando tamanhos amostrais ( $n$ ) iguais a 400, 800 e 1000, número de populações ( $k = 5$  e 10), para  $\alpha = 5\%$ . Especificou-se também que os valores de  $\pi$  no primeiro grupo fossem iguais a 0,05 e os do segundo por  $0,05 + \Delta$ .

Na Tabela 8 são apresentadas as taxas de erro tipo I por experimento sob  $H_0$  parcial em função de  $k$ ,  $n$ , e  $\Delta$  para  $\alpha = 5\%$ . Observa-se que mesmo para grandes amostras, o TCM apresentou taxas de erro tipo I significativamente ( $P < 0,01$ ) inferiores ao valor nominal adotado, independente dos valores de  $k$  e  $\Delta$ , sendo considerados conservativos.

Na Tabela 9 são apresentados os valores de poder do TCM em função de  $k$ ,  $n$  e  $\Delta$  para  $\alpha = 5\%$ . Verifica-se, de um modo geral, que os valores de poder foram pequenos em quase todas as situações, exceto para  $\Delta = 0,04$  e  $n = 800$  e 1000, e que houve uma redução expressiva do poder com o aumento de  $k = 5$  para  $k = 10$ . Também verificou-se que com o aumento do tamanho das amostras ou com o aumento dos valores de  $\Delta$ , o poder do TCM aumenta.

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 de diferenças entre os parâmetros  $\pi$  de cada grupo ( $\Delta$ ) considerando  $\pi^{(1)} = 0,05$ , para o teste de comparações múltiplas (TCM) ao nível nominal de 5%

$k$	$n$	$\Delta = 0,01$	$\Delta = 0,02$	$\Delta = 0,03$	$\Delta = 0,04$
5	400	0,86 <sup>+</sup>	0,54 <sup>+</sup>	0,73 <sup>+</sup>	0,99 <sup>+</sup>
5	800	0,72 <sup>+</sup>	0,79 <sup>+</sup>	0,90 <sup>+</sup>	1,01 <sup>+</sup>
5	1000	0,69 <sup>+</sup>	0,77 <sup>+</sup>	0,85 <sup>+</sup>	0,85 <sup>+</sup>
10	400	0,07 <sup>+</sup>	0,16 <sup>+</sup>	0,17 <sup>+</sup>	0,35 <sup>+</sup>
10	800	0,09 <sup>+</sup>	0,12 <sup>+</sup>	0,23 <sup>+</sup>	0,29 <sup>+</sup>
10	1000	0,10 <sup>+</sup>	0,06 <sup>+</sup>	0,09 <sup>+</sup>	0,29 <sup>+</sup>

<sup>+</sup> significativamente inferior ao nível nominal de significância de 1% considerando uma confiança de 99%.

Tabela 9 - Poder (%) sob  $H_0$  parcial para diferentes números de populações ( $k$ ), de tamanhos de amostras ( $n$ ) e de diferenças entre os parâmetros  $\pi$  de cada grupo ( $\Delta$ ) considerando  $\pi^{(1)} = 0,05$ , para o teste de comparações múltiplas (TCM) ao nível nominal de 5%

$k$	$n$	$\Delta = 0,01$	$\Delta = 0,02$	$\Delta = 0,03$	$\Delta = 0,04$
5	400	0,77	3,08	9,37	38,09
5	800	1,44	8,90	27,85	79,59
5	1000	1,90	12,22	38,36	90,04
10	400	0,02	0,16	0,82	7,49
10	800	0,07	0,72	4,65	37,37
10	1000	0,09	1,27	8,13	55,09

## Conclusões

O teste  $X^2$ , em geral, controlou o erro tipo I, em níveis iguais ou inferiores aos valores nominais de significância, e apresentou desempenho superior ao teste  $G^2$  que foi considerado liberal, principalmente nas situações de pequenas amostras e maiores números de populações. Na maioria da situações avaliadas, considerando pequenas amostras e diferenças entre proporções binomiais dos dois grupos pequenas e moderadas, os valores de poder do teste  $G^2$  foram superiores aos do teste  $X^2$ . A medida que essa diferença aumentou, o poder de ambos os testes se igualou e foi relativamente alto, mesmo para pequenas amostras.

O teste de comparações múltiplas assintóticas proposto para populações binomiais apresentou excelentes resultados. Como houve controle do erro tipo I, por experimento, de forma conservativa, e os resultados de poder deste teste foram bons, recomenda-se sua aplicação em situações reais. Ressaltando que, os tamanhos amostrais utilizados na experimentação agrícola estão associados a baixo poder.

NÁDIA, G. B; FERREIRA, D. F. Multiple comparisons and simultaneous test for  $k$  binomial independent parameters. *Rev. Mat. Estat.*, São Paulo, v.27, n.3, p.301-323, 2009.

- **ABSTRACT:** *A strategy used to compare several binomial proportions is the analysis of variance  $F$  test followed by tests of multiple comparisons, if the overall null hypothesis had been rejected. However, the assumptions of those tests are not satisfied in the binomial circumstance and therefore this strategy is not appropriate. Among the methods used to circumvent the problem, there are asymptotic tests proposed in the literature. This work aimed to propose an asymptotic test for multiple comparisons for binomial proportions based on quadratic forms and to evaluate their performance by means Monte Carlo simulation, and to divulge and evaluate the performance of asymptotic tests  $G^2$  and Pearson's  $X^2$  for the hypothesis of several binomial proportions. We simulated Monte Carlo samples for different configuration of  $k$  binomial populations with parameters  $\pi_i$  and  $n_i$  for the  $i$ th population,  $i = 1, 2, \dots, k$ .*

The  $X^2$  test controlled the type I error rates showing values equal to or lower than the nominal significance levels and showed superior performance than the  $G^2$  test, that was liberal, especially for the small samples and large number of populations circumstances. Since the asymptotic multiple comparisons test controlled the experimentwise type I error rates in a conservative way, and the power was high this test is recommended for real situations.

- **KEYWORDS:** Monte Carlo simulation; likelihood ratio; quadratic forms; multiple comparison procedures.

## Referências

- AGRESTI, A. *Categorical data analysis*. New York: Jonh Wiley & Sons, 2002. 734p.
- AGRESTI, A.; BINI, M.; BERTACCINI, B.; RYU, E. Simultaneous confidence intervals for comparing binomial parameters. *Biometrics*, Washington, v.64, n.4, p.1270-1275, 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 para proporções binomiais utilizando bootstrap. *Rev. Mat. Estat.*, São Paulo, v.24, n.1, p.95-112, 2006.
- BORGES, L. C.; FERREIRA, D.F. Poder e taxas de erro tipo I dos testes de Scott-Knott, Tukey e Student-Newman-Keuls sob distribuições normais e não normais dos resíduos. *Rev. Mat. Estat.*, São Paulo, v.21, n.1, p.67-83, 2003.
- DOBSON, A. J.; BARNETT, A. *An introduction to generalized linear models*. 3 ed. London: Chapman & Hall, 2008. 244p.
- FERREIRA, D.F. *Estatística multivariada*. Lavras: Editora UFLA, 2008. v.1. 662p.
- HOCHBERG, Y.; TAMHANE, A. C. *Multiple comparison procedures*. New York: Jonh Wiley & Sons, 1987. 450p.
- KRISHNAMOORTHY, K.; PENG, J. Exact properties of a new test and other tests for differences between several binomial proportions. *J. Appl. Stat. Sci.*, Hauppauge, 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. *Comp. Stat. Data Anal.*, Amsterdam, v.45, n.4, p.697-707, 2004.
- MCCULLOCH, C.E; SEARLE, S.R. *Generalized, linear, and mixed models*. New York: Jonh Wiley & Sons, 2001. 346p.
- MOOD, A.M.; GRAYBILL, F.A.; BOES, D.C. *Introduction to the theory of statistics*. 3. ed. New York: Jonh Wiley & Sons, 1974. 564 p.

- NELDER, J. A.; WEDDERBURN, R. W. M. Generalized linear models. *J. R. Stat. Soc. A*, London, v.135, p.370-384, 1972.
- PAN, W. Approximate confidence intervals for one proportion and two proportions. *Comp. Stat. Data Anal.*, Amsterdam, v.40, n.1, p.143-157, 2002.
- PIEGORSCH, W.W. Multiple comparisons for analysing dichotomous response. *Biometrics*, Washington, v.47, n.1, p.45-52, 1991.
- RAVEN, P.H.; RAY F.E.; SUSAN E.E. *Biology of plants*, 7. ed. New York: Freeman & Company, 2005. 960p.
- SANTOS, C.; FERREIRA, D.F.; BUENO FILHO, J.S.S. Novas alternativas de testes de agrupamento avaliadas por meio de simulação Monte Carlo. *Ciênc. Agrotecnol.*, Lavras, v.25, n.6, p.1382-1392, 2001.
- SCOTT, D.; DANFORD, N.; DEAN, B.; KIRKLAND, D.; RICHARDSON, C. *In vitro* chromosome aberration assays. In: DEAN, B. J. (Ed.) *Report of the UKEMS sub-committee on guidelines for mutagenicity testing*, Cambridge: Cambridge University Press, 1983. part 1, p.46-64.
- WILLIAMS, D.A. Test for differences between several small proportions. *J. R. Stat. Soc.*, London, v.37, n.3, p.421-434, 1988.

Recebido em 30.03.2009.

Aprovado após revisão em 13.08.2009.