

**INFERÊNCIAS SOBRE PROPORÇÕES BINOMIAIS:
TESTES FREQUENTISTAS E BAYESIANOS**

NÁDIA GIARETTA BIASE

2009

NÁDIA GIARETTA BIASE

**INFERÊNCIAS SOBRE PROPORÇÕES BINOMIAIS: TESTES
FREQUENTISTAS E BAYESIANOS**

Tese apresentada à Universidade Federal de Lavras como parte das exigências do Curso de Doutorado em Estatística e Experimentação Agropecuária, para a obtenção do título de “Doutor”.

Orientador

Prof. Dr. Daniel Furtado Ferreira

LAVRAS
MINAS GERAIS-BRASIL
2009

**Ficha Catalográfica Preparada pela Divisão de Processos Técnicos da
Biblioteca Central da UFLA**

Biase, Nádia Giaretta.

Inferências sobre proporções binomiais: testes frequentistas e bayesianos / Nádia Giaretta Biase. -- Lavras: UFLA, 2009.

152 p. : il.

Tese (Doutorado) - Universidade Federal de Lavras, 2009.

Orientador: Daniel Furtado Ferreira.

Bibliografia.

1. Simulação Monte Carlo. 2. Razão de verossimilhanças. 3. Formas quadráticas 4. Procedimentos de comparações múltiplas. I. Universidade Federal de Lavras. II. Título.

CDD-519.282

-519.54

NÁDIA GIARETTA BIASE

**INFERÊNCIAS SOBRE PROPORÇÕES BINOMIAIS: TESTES
FREQUENTISTAS E BAYESIANOS**

Tese apresentada à Universidade Federal de Lavras como parte das exigências do Curso de Doutorado em Estatística e Experimentação Agropecuária, para a obtenção do título de “Doutor”.

APROVADA em 23 de abril de 2009

Prof. Dr. Marcelo Tavares	UFU
Prof. Dr. Ednaldo Carvalho Guimarães	UFU
Prof. Dr. Júlio Silvío de Sousa Bueno Filho	UFLA
Profa. Dra. Thelma Sáfyadi	UFLA

Prof. Dr. Daniel Furtado Ferreira
UFLA
(Orientador)

LAVRAS
MINAS GERAIS-BRASIL

*“A esperança não é um sonho, mas uma maneira
de traduzir os sonhos em realidade.”*

(Suenens)

Agradecimentos

A Deus, que sempre iluminou e guiou a minha vida, dando-me forças para vencer e concretizar meus sonhos.

A minha querida mãe, pela paciência e incondicional amor, que nunca mediu esforços para que meus objetivos fossem alcançados e sempre apoiou em todas as decisões de minha vida.

Ao meu querido pai, pela confiança e infinito amor, que sempre lutou incansavelmente pela minha educação e foi o meu suporte para vencer todos os obstáculos encontrados.

As minhas irmãs, Érica e Adriele, pelo carinho, atenção e pelo estímulo constante no decorrer destes anos, para que eu vencesse minhas apreensões. Sei que vocês e nossos pais almejavam tanto quanto eu a conquista desse projeto de vida.

Ao meu cunhado, Edivânio, pelo carinho, consideração e amizade.

Ao meu orientador, Daniel Furtado Ferreira, pelo profissionalismo, competência intelectual, disponibilidade e entusiasmo com que me orientou e, principalmente, pela amizade, paciência e por todos os ensinamentos e conselhos transmitidos durante esses anos, não restritos somente à vida acadêmica.

À Universidade Federal de Lavras, pela oportunidade de realizar o doutorado. Em especial, ao Departamento de Ciências Exatas (DEX), por todas as condições oferecidas para minha formação, aos funcionários, pela acolhida simples e carinhosa e aos professores, pela amizade e por ajudarem a enriquecer os meus conhecimentos.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes), pelo apoio financeiro concedido durante o período de realização deste trabalho.

Aos membros da banca examinadora, pelas sugestões e contribuições para o aprimoramento desse trabalho.

À amiga de longa data, Maria Imaculada, pelo carinho, amizade, troca de conhecimentos, apoio e por todos os momentos inesquecíveis durante essa caminhada.

Às amigas Patrícia, Verônica, Andréa e Fabrícia pela ajuda e a amizade com que sempre pude contar e também às amigas Gabriella, Flávia, Franciella, Jessica e Marília que compartilharam harmoniosamente comigo vários momentos do cotidiano. Nunca me esquecerei de tudo o que todas vocês fizeram por mim.

A todos os colegas da Pós-Graduação em Estatística que conviveram comigo em Lavras, por todas as palavras e gestos de amizade manifestado durante esses anos.

A todos os amigos do Grupo de Partilha de Profissionais (GPP), pelas orações e momentos agradáveis de confraternização.

Finalmente, a todos aqueles que, de alguma forma, contribuíram e acreditaram no meu crescimento e sucesso. Deus os abençoe!!!

SUMÁRIO

LISTA DE TABELAS	i
LISTA DE FIGURAS	vii
RESUMO	viii
ABSTRACT	ix
CAPÍTULO 1	1
1 Introdução Geral	2
2 Referencial Teórico	5
2.1 Procedimentos de comparações múltiplas (PCM)	5
2.2 Erros envolvidos e poder	8
2.3 Pressuposições dos procedimentos de comparações múltiplas	12
2.4 Testes de comparações múltiplas	14
2.4.1 Procedimentos de comparações múltiplas de proporções binomiais	14
2.4.2 Teste de comparações múltiplas via <i>bootstrap</i>	17
2.4.3 Teste de comparações múltiplas bayesianos	20
2.5 Inferência bayesiana	22
2.5.1 Distribuição <i>a priori</i>	25
2.5.2 Métodos de simulação Monte Carlo via Cadeias de Markov (MCMC)	26
2.5.3 Estimação usando a inferência bayesiana	28
2.6 Teste de razão de verossimilhanças	30
2.6.1 Estimação	32
2.7 Inferências sobre proporções multinomiais	35
2.8 Testes de distribuições baseadas em formas quadráticas	38
2.9 Aplicações testes assintóticos	41
2.10 Simulação Monte Carlo	43
3 Referências Bibliográficas	45

CAPÍTULO 2: Comparações múltiplas e teste simultâneo para parâmetros binomiais independentes de k populações	53
1 RESUMO	54
2 ABSTRACT	55
3 Introdução	56
4 Metodologia	58
4.1 Teste de razão de verossimilhanças	58
4.2 Teste assintótico qui-quadrado	61
4.3 Comparações múltiplas	65
4.4 Simulações Monte Carlo	65
5 Resultados e Discussão	69
5.1 Testes para a hipótese nula global	69
5.1.1 Erro tipo I sob H_0 completa	69
5.1.2 Poder sob H_0 parcial	75
5.2 Comparações múltiplas	83
5.2.1 Erro tipo I sob H_0 completa	84
5.2.2 Erro tipo I sob H_0 parcial	87
5.2.3 Poder sob H_0 parcial	91
6 Conclusões	98
7 Referências Bibliográficas	99
CAPÍTULO 3: Inferência Bayesiana para k populações binomiais independentes	101
1 RESUMO	102
2 ABSTRACT	103
3 Introdução	104
4 Metodologia	106
4.1 Teste bayesiano para igualdade de proporções binomiais	106
4.2 Comparações múltiplas bayesianas	112

4.3 Simulação Monte Carlo	113
5 Resultados e Discussão	116
5.1 Teste bayesiano	116
5.2 Erro tipo I	116
5.2.1 Erro tipo I sob H_0 completa	116
5.3 Poder	119
5.3.1 Poder sob H_0 parcial	119
5.4 Teste de comparações múltiplas bayesiano	123
5.5 Erro tipo I	124
5.5.1 Erro tipo I sob H_0 completa	124
5.5.2 Erro tipo I sob H_0 parcial	126
5.6 Poder	130
5.6.1 Poder sob H_0 parcial	130
6 Conclusões	134
7 Referências Bibliográficas	135
ANEXOS	137

LISTA DE TABELAS

1.1	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 os testes de <i>bootstrap</i> de Pan (Pan) e de máxima verossimilhança (MV), no valor nominal de 5%.	18
1.2	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 os testes de razão de verossimilhanças (G^2) e formas quadráticas (X^2), ao nível nominal de 5%.	70
1.3	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 os testes de razão de verossimilhanças (G^2) e formas quadráticas (X^2), ao nível nominal de 1%.	72
1.4	Poder (%), sob H_0 parcial, para diferentes números de populações (k), de tamanhos de amostras (n) e de diferenças entre as proporções binomiais (Δ), para os testes de razão verossimilhanças (G^2) e formas quadráticas (X^2), ao nível nominal de 5%.	76
1.5	Poder (%), sob H_0 parcial, para diferentes números de populações (k), de tamanhos de amostras (n) e de diferenças entre as proporções binomiais (Δ), para os testes de razão verossimilhanças (G^2) e formas quadráticas (X^2), ao nível nominal de 1%.	77
1.6	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 os testes de razão verossimilhanças (G^2) e formas quadráticas (X^2), ao nível nominal de 5%.	79

1.7	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 os testes de razão verossimilhanças (G^2) e formas quadráticas (X^2), ao nível nominal de 1%.	80
1.8	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 π para o teste de comparações múltiplas (TCM), ao nível nominal de 5%.	85
1.9	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 π para o teste de comparações múltiplas (TCM), ao nível nominal de 1%.	87
1.10	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 (TCM), ao nível nominal de 5%.	88
1.11	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 (TCM), ao nível nominal de 1%.	89
1.12	Taxas de erro tipo I por experimento (%), 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 de comparações múltiplas (TCM), ao nível nominal de 5%.	91

1.13 Poder (%), sob H_0 parcial, para diferentes números de populações (k), tamanhos de amostras (n) e diferenças entre os parâmetros π de cada grupo (Δ), para o teste de comparações múltiplas (TCM), ao nível nominal de 5%.	92
1.14 Poder (%), sob H_0 parcial, para diferentes números de populações (k), tamanhos de amostras (n) e diferenças entre os parâmetros π de cada grupo (Δ), para o teste de comparações múltiplas (TCM), ao nível nominal de 1%.	93
1.15 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 de comparações múltiplas (TCM), ao nível nominal de 5%.	95
1.16 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 π de cada grupo (Δ) considerando $\pi^{(1)} = 0,05$, para o teste de comparações múltiplas (TCM), ao nível nominal de 5%.	97
1.17 Poder (%), sob H_0 parcial, para diferentes números de populações (k), de tamanhos de amostras (n) e de diferenças entre os parâmetros π de cada grupo (Δ) considerando $\pi^{(1)} = 0,05$, para o teste de comparações múltiplas (TCM), ao nível nominal de 5%.	97
2.18 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%.	117

2.19	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%.	118
2.20	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%.	120
2.21	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%.	121
2.22	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%.	123
2.23	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%.	125

2.24	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%.	126
2.25	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%.	127
2.26	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%.	128
2.27	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%.	129
2.28	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%.	131

2.29 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 1%. 132

LISTA DE FIGURAS

1.1	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.	73
1.2	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 $\alpha = 1\%$, considerando a hipótese H_0 completa.	74
1.3	Poder, sob H_0 parcial, dos testes G^2 e X^2 , em função da diferença Δ , com $k = 5$, $n = 10$ e valores nominais de significância iguais (a) $\alpha = 5\%$ e (b) $\alpha = 1\%$ para o teste G^2 e (c) $\alpha = 5\%$ e (d) $\alpha = 1\%$, para o teste X^2	82

RESUMO

BIASE, Nádia Giaretta. **Inferências sobre proporções binomiais:** testes frequentistas e bayesianos. 2009. 152p. Tese (Doutorado em Estatística e Experimentação Agropecuária) - Universidade Federal de Lavras, Lavras, MG. *

Inferências sobre várias proporções binomiais são frequentemente realizadas por meio da análise de variância e dos procedimentos de comparações múltiplas convencionais. O problema é que, nesse caso, algumas pressuposições dos testes aplicados são violadas e, assim, as inferências são não confiáveis. Uma solução para minimizar o problema consiste em aplicar testes assintóticos e procedimentos bayesianos. O presente trabalho teve por objetivos propor testes de comparações múltiplas clássicos e bayesianos, bem como propor um teste bayesiano para a hipótese global de igualdade de várias proporções binomiais e divulgar os testes assintóticos G^2 e X^2 de Pearson no contexto das distribuições binomiais. O desempenho de todos os testes propostos foi avaliado por meio de simulação Monte Carlo. As taxas de erro tipo I do teste bayesiano proposto e dos testes assintóticos G^2 e X^2 de Pearson, para testar a hipótese de igualdade de várias proporções binomiais, foram avaliadas sob H_0 completa e as taxas de erro tipo I, por experimento dos testes de comparações múltiplas sob H_0 completa e parcial. Os valores de poder de todos os testes foram mensurados apenas para H_0 parcial. Foram geradas k populações binomiais independentes com parâmetros π_i e n_i , $i = 1, 2, \dots, k$, e simuladas amostras de Monte Carlo para cada configuração envolvendo as combinações das quantidades k , n_i 's e π_i 's. Para os testes bayesianos, consideraram-se *prioris* conjugadas betas com parâmetros α_i e β_i . As simulações sob H_0 completa foram realizadas considerando π_i 's idênticos e fixados em 0,01, 0,1 e 0,5, com populações $k = 2, 5$ e 10 e tamanhos amostrais $n_i = 10, 30$ e 100, para cada população e, para H_0 parcial, considerou-se uma diferença Δ entre os valores de π de dois grupos distintos, variando no intervalo de 0,01 a 0,9. O teste bayesiano e o teste X^2 de Pearson para a igualdade de várias proporções binomiais apresentaram excelentes performances, controlando o erro tipo I em praticamente todas as situações, em níveis iguais ou inferiores aos valores nominais. O teste G^2 é liberal, nos casos de pequenas amostras e maiores números de populações. Os valores de poder destes testes são relativamente altos, principalmente se as diferenças entre as proporções binomiais dos dois grupos são grandes. Os testes de comparações múltiplas clássicos e bayesianos para populações binomiais apresentaram excelentes resultados e, em geral, são conservativos.

* **Orientador:** Daniel Furtado Ferreira - UFLA

ABSTRACT

BIASE, Nádia Giaretta. **Inferences about binomial proportions:** frequentist and bayesian tests. 2009. 152p. Thesis (Doctor in Statistics and Agricultural Experimentation) - Federal University of Lavras, Lavras, MG.*

Inferences about several binomial proportions are often made by means of an analysis of variance followed by conventional multiple comparisons procedures. The problem of this approach is that some tests assumptions are violated and thus the inferences are not reliable. A solution to minimize the problem is to apply asymptotic and bayesian tests. This work aimed to propose tests classical and bayesian of multiple comparisons and a bayesian test for the overall hypothesis of equality of several binomial proportions and to divulge the asymptotic Pearson's X^2 and G^2 . The performance of all proposed tests were evaluated by means of Monte Carlo simulation. The type I error rates of the bayesian proposed test and of the asymptotic tests Pearson's X^2 and G^2 for the overall hypothesis of equality of several binomial proportions were evaluated under complete H_0 and the experimentwise type I error rates and also for the multiple comparisons procedures under complete and partial H_0 . The power of all tests were computed only for partial H_0 . Independent binomial populations with parameters π_i and n_i were sampled, $i = 1, 2, \dots, k$ considering different configurations involving combinations of k , n_i 's and π_i 's. For the bayesian tests ones considered conjugated betas prior with parameters α_i and β_i . The simulations under complete H_0 were made considering identical π_i 's settled in 0.01, 0.1 and 0.5, with number of populations $k = 2, 5$ and 10 and sample sizes $n_i = 10, 30$ and 100 for each population. For partial H_0 it was considered a difference Δ between the values of π of two different groups, varying in the interval 0.01 to 0.9. The bayesian and Pearson's X^2 tests for equality of several binomial proportions showed excellent performance, controlling the type I error rates in almost all cases, at levels below or equal to the nominal levels. The G^2 test was liberal in cases of small samples and large number of populations. The powers of the tests are relatively high, especially if the differences of the binomial proportions between the two groups are large. The classical and bayesian multiple comparisons tests for the binomial proportions showed excellent performance and in general are conservatives.

***Guidance Committee:** Daniel Furtado Ferreira - UFLA

CAPÍTULO 1

1 Introdução Geral

Nas pesquisas científicas, é de interesse de muitos pesquisadores inferir sobre duas ou mais médias populacionais. Para testar a hipótese de igualdade das médias populacionais, frequentemente aplica-se o teste F em uma análise de variância e, quando essa hipótese é rejeitada, vários procedimentos de comparações múltiplas são utilizados para investigar as diferenças entre as proporções das populações, tais como os testes Tukey, Student-Newman-Keuls (SNK), Duncan e Scheffé.

Para garantir que os resultados da aplicação desses testes sejam válidos, as pressuposições de independência das observações, normalidade dos resíduos e homogeneidade das variâncias devem ser satisfeitas. Particularmente no caso da independência das observações, há o controle, pelo pesquisador, por meio da utilização de casualização. As demais pressuposições devem ser checadas caso a caso.

Quando o problema de comparar duas ou mais proporções binomiais nas pesquisas está em foco, alguns pesquisadores fazem uso da análise de variância, teste F e dos procedimentos de comparações múltiplas convencionais. No entanto, esse procedimento constitui apenas uma aproximação, uma vez que a pressuposição de normalidade dos resíduos e de homogeneidade das variâncias é violada quase que certamente. Assim, as inferências são não confiáveis, embora existam alternativas para contornar ou minimizar o problema.

Técnicas estatísticas mais apropriadas, por considerarem a distribuição específica dos dados, como a binomial, podem ser utilizadas, como é o caso dos modelos lineares generalizados. Nesse caso, a modelagem é feita pela família exponencial que abrange o modelo binomial. O problema dessa abordagem é que as distribuições das estatísticas dos testes da hipótese nula global de igualdade das proporções das diferentes populações binomiais são assintóticas e, portanto, a inferência pode ter baixa qualidade em pequenas amostras. Ademais, não existem procedimentos de comparações múltiplas

entre os testes relacionados à família exponencial.

Outra opção é utilizar os métodos computacionalmente intensivos, como os testes *bootstrap* e de permutação. Não são encontrados relatos, na literatura, para o teste global de igualdade das proporções binomiais, mais existem relatos do seu uso em comparações múltiplas. Para esse fim, os resultados são de alta qualidade. O problema é a necessidade de implementação de rotinas para a realização desses testes, o que é uma tarefa impeditiva para a maioria dos pesquisadores. Além disso, os programas comerciais de análise estatística, ou até mesmo os gratuitos, não têm opções ou funções para realizar essa tarefa. Outro problema que, em épocas mais antigas, era relevante é o grande número de reamostragens necessárias para a aplicação do teste, embora isso não seja mais relevante, considerando a capacidade de processamento dos computadores atuais.

Para o teste global da igualdade de várias proporções binomiais existem os testes assintóticos da razão de verossimilhanças (estatística G^2) e o teste X^2 de Pearson. Esses testes são pouco conhecidos para essa finalidade, pois são, em geral, utilizados nos modelos multinomiais e nas tabelas de contingência. Estudos de desempenho desses testes foram encontrados no trabalho de Williams (1988) e Krishnamoorthy & Peng (2008), embora Krishnamoorthy et al. (2004) relatem o emprego de testes para igualdade de várias proporções binomiais com um valor de referência específico. Entretanto, nenhum trabalho foi encontrado, envolvendo aproximações assintóticas, relatando o uso de testes de comparações múltiplas para proporções binomiais.

A inferência bayesiana tem tido um grande papel nos trabalhos científicos. O grande apelo do uso das técnicas de inferência nessa área da estatística é a possibilidade de incorporar o conhecimento *a priori* do pesquisador sobre os parâmetros. No caso particular de realizações de inferências sobre várias proporções binomiais, principalmente no caso das comparações múltiplas, nenhum relato foi encontrado na literatura científica.

Por todas essas razões, o presente trabalho foi realizado com o objetivo de pro-

por um teste bayesiano para a hipótese global de igualdade de várias proporções binomiais e divulgar os testes assintóticos G^2 e X^2 de Pearson, no contexto das distribuições binomiais, bem como propor testes de comparações múltiplas clássicos e bayesianos, com a avaliação do desempenho por simulação Monte Carlo.

2 Referencial Teórico

2.1 Procedimentos de comparações múltiplas (PCM)

Os experimentos, na pesquisa científica, são planejados com a finalidade de propiciar a comparação de médias de diferentes níveis do fator. No entanto, para verificar se existem diferenças reais entre os níveis de qualquer fator de efeitos fixos, um teste de hipótese pode ser formulado. A hipótese de nulidade global (H_0) de interesse é:

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_i = \dots = \mu_k \quad (2.1)$$

que estabelece não existirem diferenças entre os k níveis do fator e deve ser testada utilizando-se o teste F (Machado et al., 2005).

Quando o pesquisador decide por não rejeitar a hipótese de nulidade global, admite-se a não existência do efeito do fator sobre a variável resposta. Caso contrário, se essa hipótese é rejeitada, adota-se a hipótese alternativa, representada por H_1 ou H_a como verdadeira, que supõe existir pelo menos uma diferença entre os níveis do fator. Surge, então, a necessidade de averiguar a que se devem as diferenças e quais são os níveis do fator que diferem entre si (Hochberg & Tamhane, 1987).

Para investigar as diferenças específicas entre níveis do fator ou combinações lineares de médias do fator, várias técnicas estatísticas podem ser empregadas. Dentre elas têm-se os procedimentos de comparações múltiplas (Rafter et al., 2002). Machado et al. (2005) afirmam que a aplicação adequada desses procedimentos está relacionada com os seguintes tipos de níveis do fator (tratamento) em estudo:

- i) se os níveis do fator são quantitativos, é aconselhável utilizar os métodos de análise de regressão;
- ii) se os níveis do fator são qualitativos com uma estruturação que propõe com-

parações pré-planejadas entre os níveis do fator, recomenda-se, primeiramente, aplicar contrastes e, depois, um teste específico;

- iii) se os níveis do fator são qualitativos e não estruturados, procedimentos de comparações múltiplas (PCM) são indicados.

A comparação de médias dos níveis do fator é planejada quando ela é definida *a priori* ou durante a fase de planejamento do experimento. A decisão a respeito da possibilidade de as comparações serem planejadas ou não depende do tipo de fator em estudo e dos objetivos do experimento. Uma comparação não é planejada quando ela é definida em função daquilo que foi observado após a realização do experimento (Machado et al., 2005)

Segundo Hsu (1996), os PCM são utilizados para comparar duas ou mais médias e, apesar de serem usados consistentemente na pesquisa científica, muitas vezes são empregados incorretamente. De acordo com Hinkelmann & Kempthorne (1987), em várias situações experimentais, o pesquisador tem interesse em realizar um grande número de comparações, como, por exemplo, comparações sugeridas pelos dados ou todas as possíveis combinações entre duas delas. Nesse caso, algumas precauções devem ser tomadas para que os procedimentos de inferência sejam empregados corretamente quando se realizam testes de hipóteses ou estimação, caso contrário, a falta de transitividade da não significância pode conduzir a muitos resultados significativos.

Gopalan & Berry (1998) comentam que a falta de transitividade da não significância é um dos problemas mais difíceis enfrentados por estatísticos e outros pesquisadores e que os PCM são casos especiais dessa falta de transitividade.

Machado et al. (2005) ressaltam que são muitos os problemas envolvidos na aplicação dos PCM e mencionam alguns desses problemas:

1. grande número de parâmetros a serem considerados: são realizadas todas as comparações duas a duas, conseqüentemente, o número de parâmetros cresce muito

com o aumento do número de níveis k do fator em estudo;

2. falta de transitividade da não-significância: conhecido por ambiguidade dos resultados. Como ilustração, considere três médias; a maior pode diferir da menor, mas ambas não diferirem da média intermediária;
3. dificuldade de interpretação: devido à falta de transitividade ou ambiguidade dos resultados;
4. teste global não significativo: o teste F da hipótese de igualdade dos k efeitos do fator apresentam resultados significativos em um nível nominal α , entretanto, os testes de comparações múltiplas apresentam resultados não significativos nesse mesmo nível de significância nominal, indicando que os níveis dos fatores são iguais.

Assim, a seleção do método apropriado de comparações múltiplas pelo qual se deve optar depende das qualidades estatísticas desse procedimento. Essa qualidade está relacionada com o tipo de erro que é controlado e da forma como esses erros são controlados (Machado et al., 2005). Para aplicação da maioria dos PCM, uma padronização das estatísticas utilizadas precisa ser realizada por meio do estimador do erro padrão da diferença das médias entre dois níveis do fator (O'Neill & Wetherill, 1971; Perry, 1986). Esse estimador envolve o quadrado médio do erro obtido na análise de variância e é conhecido por diferença mínima significativa (DMS). A expressão generalizada da DMS é dada por:

$$DMS = \gamma S_d,$$

em que: $S_d = \sqrt{2QME/r}$ é o estimador do erro padrão da diferença de duas médias em um delineamento balanceado; QME é o quadrado médio do erro da análise de variância correspondente a ν graus de liberdade; r é o número de repetições das médias a serem comparadas e γ depende do método, dos graus de liberdade do erro e do número

de comparações simultâneas.

Independentemente do PCM utilizado, a diferença observada entre duas médias, consecutivas ou não, é confrontada com um valor crítico apropriado. Se o valor absoluto da diferença observada entre duas médias exceder o valor crítico, ela é considerada significativa e, conseqüentemente, as médias são diferentes. Caso contrário, a diferença é considerada não-significativa e, portanto, as médias são estatisticamente iguais. Os valores críticos alternam de um procedimento para outro. Com isso, PCM distintos, aplicados a um mesmo conjunto de dados, podem apresentar diferentes resultados (Carmar & Swanson, 1973).

Quando estudos de simulações são realizados para avaliar o desempenho dos PCM, pesquisadores frequentemente geram dados experimentais sob a situação de nulidade parcial ou H_0 parcial, para retratar situações em que determinados níveis do fator em estudo são diferentes e outros níveis são considerados iguais. Nessas situações, a hipótese de nulidade global (2.1) é que é testada ao serem aplicados os testes. Isso permite aos pesquisadores reproduzirem situações reais que são consideradas comuns (Silva et al., 1999; Santos et al., 2001).

2.2 Erros envolvidos e poder

Ao realizar testes de hipóteses, o pesquisador deve levar em consideração o risco global de tomar uma decisão errada. Ao tomar a decisão de rejeitar ou não a hipótese de nulidade, o pesquisador está sujeito a incorrer em três possíveis erros (Mood et al., 1974; Rafter et al., 2002; Steel & Torrie, 1980), descritos a seguir.

O erro tipo I é cometido quando a hipótese nula verdadeira, da forma $H_0 : \mu_i = \mu_{i'}, i \neq i' = 1, 2, \dots, k$, é rejeitada. Esse erro é controlado diretamente pelo pesquisador e a probabilidade de se cometer esse erro é representada pelo nível de significância α , dada por: $P[\text{Erro Tipo I}] = P[\text{rejeitar } H_0 | H_0 \text{ verdadeira}] = \alpha$.

O erro tipo II é cometido quando a hipótese nula falsa não é rejeitada. Esse erro não é controlado diretamente pelo pesquisador e a probabilidade (β) de se cometer esse erro é dada por: $P[\text{Erro Tipo II}] = P[\text{não rejeitar } H_0 | H_0 \text{ falsa}] = \beta$.

O último tipo de erro, conhecido como erro tipo III, é cometido quando uma hipótese nula falsa é rejeitada a favor da hipótese alternativa errada. Nesse caso, afirma-se que uma média é superior a outra, quando ocorre exatamente o contrário (Ramalho et al., 2000; Machado et al., 2005). A probabilidade associada a esse erro é: $P[\text{Erro Tipo III}] = P[\text{rejeitar } H_0, \text{ mas a favor de } H_1 \text{ errada} | H_0 \text{ falsa}]$.

O poder do teste é a probabilidade ($1 - \beta$) de rejeitar a hipótese nula H_0 , quando ela realmente é falsa, ou seja, é a capacidade do teste em determinar todas as reais diferenças entre os níveis do fator (Ramalho et al., 2000).

O desempenho da maioria dos testes de comparações múltiplas é avaliado em relação ao controle da taxa de erro tipo I, para assegurar o nível de probabilidade pretendido em um conjunto de várias comparações. As probabilidades de se cometerem os erros tipo I e II são inversamente proporcionais. Portanto, é necessário manter um certo equilíbrio, pois, ao controlar de maneira excessiva a taxa de erro tipo I, a taxa de erro tipo II aumenta e o poder do teste diminui, levando a aceitar como iguais médias diferentes. Quando isso ocorre, o teste é considerado conservativo (Carmer & Swanson, 1973). Esse fato pode causar um efeito extremamente indesejável, uma vez que um dos objetivos dos experimentos é discriminar tratamentos por meio de suas médias (Machado et al., 2005).

Por outro lado, se a taxa de erro tipo I for elevada (superior ao valor nominal de significância α), a taxa de erro tipo II diminui e o poder do teste aumenta. Nesse caso, o teste é considerado poderoso e, ao mesmo tempo, liberal (Hochberg & Tamhane, 1987).

Teoricamente, algumas medidas podem ser tomadas para minimizar a taxa de erro tipo II. Uma medida está relacionada com a escolha apropriada do teste e da avaliação criteriosa das suas pressuposições, que devem ser atendidas. Se essas condições

forem satisfeitas com êxito, existe uma garantia de maior poder. Uma segunda medida que pode possibilitar o teste ter o maior poder possível é a determinação do tamanho da amostra, desde que não aumente demasiadamente o custo da pesquisa a ser realizada. Por fim, a fixação do nível de significância α entre 0,10 e 0,01 é, sempre que possível, uma atitude recomendável, por causa da relação inversa entre as taxas de erro tipo I e II (Ferreira, 2005).

Outra questão que merece ser enfatizada refere-se ao desempenho das taxas de erro tipo I e II, resultantes da aplicação de um único teste, que são diferentes daquelas decorrentes da aplicação de uma sequência de k testes. Se a probabilidade de cometer o erro tipo I for igual a α , $(1 - \alpha)$ é a probabilidade de que, em um destes testes, não seja cometido o erro tipo I. Então, se forem realizados k testes independentes, a probabilidade de não cometer o erro tipo I em nenhum dos testes é $(1 - \alpha)^k$ e $1 - (1 - \alpha)^k$ é a probabilidade máxima de se cometer o erro tipo I em pelo menos um dos k testes. Como consequência desse fato, várias estratégias surgiram para garantir a taxa de erro tipo I para todas as comparações. Procedimentos de inferência que asseguram uma probabilidade conjunta $(1 - \alpha)$ contra o erro tipo I são denominados procedimentos de inferência simultânea (Machado et al., 2005).

Quanto às taxas de erro tipo III, Carmer & Swanson (1973) mostraram que elas são baixas e, geralmente, desprezíveis e que, normalmente, são medidas considerando todas as comparações (*comparisonwise*) ou por experimento (*experimentwise*).

A escolha do teste de comparações múltiplas a ser aplicado deve levar em consideração o controle desses erros. Inúmeros trabalhos comparam os PCM em relação às taxas de erro tipo I por meio de simulação computacional, pois executar analiticamente essa tarefa é bastante complicado (Carmer & Swanson, 1973; Percin & Barbosa, 1988; Borges & Ferreira, 2003).

Segundo Steel & Torrie (1980), existem duas maneiras básicas de calcular as taxas de erro tipo I nos PCM. A primeira maneira consiste em medir a taxa de erro tipo

I por comparação, conhecida por *comparisonwise* ou *per-comparison error rate* (TPC) e refere-se à probabilidade de se rejeitar uma hipótese verdadeira em todas as possíveis combinações de médias de níveis do fator, tomadas duas a duas:

$$\text{TPC} = \frac{\text{Número de inferências erradas}}{\text{Número total de inferências}}.$$

A segunda maneira consiste em calcular a taxa de erro tipo I por experimento, denominada de *experimentwise error rate* (TPE), definida como sendo a probabilidade de se realizar pelo menos uma inferência errada por experimento:

$$\text{TPE} = \frac{\text{Número experimentos com pelo menos uma inferência errada}}{\text{Número total de experimentos}}.$$

De acordo com Gopalan & Berry (1998), muitos PCM controlam apenas uma dessas duas taxas de erro tipo I. Além disso, ao tentar controlar a taxa de erro tipo I por experimento, quando o número de comparações múltiplas a serem realizadas é grande, os pesquisadores se deparam com problemas relacionados à redução considerável do poder em detectar diferenças significativas entre os níveis do fator. Benjamini & Hochberg (1995) apresentaram um critério alternativo de proteção do erro tipo I por experimento para solucionar esse problema. Esse critério é denominado *false discovery rate* (FDR) e corresponde à proporção esperada de erro tipo I cometido entre todas as hipóteses nulas rejeitadas. Segundo esses autores, quando um grande número de hipóteses é testado, pode ser mais interessante controlar a proporção de falsas rejeições do que manter uma baixa probabilidade de, pelo menos, uma falsa rejeição.

O critério FDR é menos conservativo do que o TPE e, por essa razão, apresenta valores de poder mais elevados. A implementação desse critério é simples e pode ser aplicada tanto para dados independentes como também para dados dependentes. Vários trabalhos avaliaram o desempenho desse critério. Dentre eles têm-se Benjamini & Liu (1999), Storey (2002) e Tsai et al. (2003). Storey (2002) estimou as FDRs fixando a

região de rejeição e verificou que esse método apresenta grande aplicabilidade, precisão e alto poder. Tsai et al. (2003) avaliaram a FDR utilizando *bootstrap* por meio de simulação Monte Carlo e concluíram que este procedimento apresenta bons desempenhos.

2.3 Pressuposições dos procedimentos de comparações múltiplas

Ao realizar inferência, quase todos os procedimentos apresentam algumas pressuposições que devem ser satisfeitas pelos dados para garantir que os resultados sejam válidos (Rafter et al., 2002). Assim, a aplicação dos PCM deve satisfazer a algumas pressuposições que são as mesmas estabelecidas para a aplicação do teste F na análise de variância. De acordo com Hochberg & Tamhane (1987), essas pressuposições são:

- aditividade: os efeitos dos fatores que ocorrem no modelo matemático devem ser aditivos;
- independência das observações: os erros ou desvios, devidos ao efeito de fatores não controlados, não devem ser correlacionados;
- normalidade dos resíduos: os erros ou desvios, devidos ao efeito de fatores não controlados, devem possuir uma distribuição normal;
- homogeneidade das variâncias: os erros ou desvios, devidos ao efeito de fatores não controlados, devem possuir variância comum.

Rafter et al. (2002) afirmam que qualquer uma dessas pressuposições pode ser violada para um determinado conjunto de dados, no entanto, é importante ter a consciência de que essa violação pode causar impactos nas inferências. Segundo estes autores, um procedimento que é insensível à violação de uma ou mais dessas pressuposições é considerado robusto. A pressuposição da independência das observações é a menos provável de ser violada porque está sob controle do pesquisador.

Quando os PCM não satisfazem às pressuposições de normalidade e homogeneidade de variâncias, existem indicações de que muitos deles não são considerados um procedimento robusto (Hochberg & Tamhane, 1987). Segundo Rafter et al. (2002), geralmente, quando a pressuposição de normalidade é violada moderadamente, a taxa de erro tipo I por experimento de muitos PCM é ligeiramente maior que o valor nominal adotado. Ringland (1983) acrescenta, ainda, que, mesmo que a diferença entre a taxa de erro tipo I e o valor nominal não seja tão discrepante ao aplicar um único teste de comparação múltipla, ao realizar inferências múltiplas, essa diferença aumenta proporcionalmente ao número de comparações realizadas.

Vários métodos alternativos podem ser utilizados para contornar as dificuldades de aplicação dos PCM nas situações de não normalidade ou de variâncias heterogêneas sob modelos probabilísticos normais ou não normais (Machado et al., 2005). Nessas situações, Steel & Torrie (1980) ressaltam que métodos de análise mais robustos podem ser aplicados, como, por exemplo, os modelos lineares generalizados e os métodos não-paramétricos.

A técnica dos modelos lineares generalizados (MLGs), apresentada por Nelder & Wedderburn (1972), é uma extensão dos modelos lineares clássicos e permite que a distribuição da variável resposta seja normal ou não normal, podendo ser qualquer distribuição da família exponencial. Essa técnica proporciona também maior flexibilidade para a relação funcional entre a média da variável resposta e o preditor linear.

Os métodos não-paramétricos são procedimentos da inferência estatística que não fazem qualquer suposição explícita sobre a forma da distribuição dos dados, tendo, portanto, menores exigências para a sua aplicação. A base dos testes não-paramétricos está na ordenação (postos) dos dados e não em seu valor intrínseco e na aleatorização, em que se consideram todas as possíveis permutações (rearranjos) dos dados (Hochberg & Tamhane, 1987).

Machado et al. (2005) dão ênfase à aplicação dos procedimentos robustos de comparações múltiplas baseados na reamostragem *bootstrap*. O método de *bootstrap* é um processo de reamostragem das amostras que permite obter as estimativas dos parâmetros sem pressupor a distribuição dos estimadores. Esse método tem apresentado resultados satisfatórios e tem sido utilizado por muitos pesquisadores. Recentemente, uma outra abordagem estatística que vem sendo utilizada para solucionar os problemas envolvendo comparações múltiplas é a inferência bayesiana, que descreve toda quantidade desconhecida por meio de probabilidades. Estudos envolvendo comparações múltiplas via *bootstrap* e inferência bayesiana são apresentados, respectivamente, nos tópicos 2.4.2 e 2.4.3.

2.4 Testes de comparações múltiplas

2.4.1 Procedimentos de comparações múltiplas de proporções binomiais

Piegorsch (1991) considerou vários procedimentos simultâneos para realizar comparações múltiplas entre proporções binomiais, dando maior ênfase à construção de intervalos simultâneos para várias funções da probabilidade de sucesso π_i , $i = 1, 2, \dots, k$. As inferências basearam-se na normalidade assintótica do estimador de máxima verossimilhança de π_i . Aplicações específicas, incluindo todos os pares de comparações ($\pi_i - \pi_{i'}, i \neq i' = 1, 2, \dots, k$) e comparações com um tratamento controle ($\pi_i - \pi_1, i \neq 1$), foram avaliadas.

Primeiramente, ele considerou intervalos simultâneos para qualquer conjunto finito de contrastes, utilizando a aproximação de Bonferroni aplicada ao intervalo de confiança de Wald, dados por: $\sum_{i=1}^k \ell_{hi} \hat{\pi}_i \pm \Phi^{-1}(1 - \frac{\alpha}{2G})(\sum_{i=1}^k \ell_{hi} \nu_i)^{\frac{1}{2}}$ para a h -ésima combinação linear das proporções, $h=1, 2, \dots, G$, em que $\Phi(\cdot)$ é a distribuição normal acumulada, $G = k(k - 1)/2$ e ν_i é a variância estimada ($\nu_i = \frac{\hat{\pi}_i(1-\hat{\pi}_i)}{n_i}$). Em uma segunda etapa, o autor considerou um método implementado por Hochberg & Tamhane

(1987), usando o intervalo de Wald juntamente com a distribuição da amplitude normal padronizada, para comparar todas as diferenças $\pi_i - \pi_{i'}, i \neq i'$, por meio do intervalo: $\hat{\pi}_i - \hat{\pi}_{i'} \pm \frac{1}{\sqrt{2}} Q_{k,\infty}(\alpha)(\nu_i + \nu_{i'})^{1/2}, \forall i < i'$, em que $Q_{k,\infty}(\alpha)$ é o quantil superior $100(\alpha)\%$ da distribuição da amplitude normal padronizada. Esses dois métodos apresentaram pobres desempenhos, tendo taxas de erro consideravelmente maiores do que o valor nominal quando os tamanhos amostrais eram pequenos.

Piegorsch (1991) mostrou também que um melhor desempenho desses intervalos foi obtido utilizando-se um procedimento de estimação intervalar simultâneo, 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 $\pi_i - \pi_{i'}$ é: $(1 - d^2)^{-1}[(\hat{\pi}_i - \hat{\pi}_{i'}) \pm d\{(2 - \tilde{\theta}_{ij})\tilde{\theta}_{ij}(1 + d^2 - (\hat{\pi}_i - \hat{\pi}_{i'})^2)\}^{1/2}]$, em que $\tilde{\theta}_{ij} = \{n(\hat{\pi}_i - \hat{\pi}_{i'}) + 1\}/(n + 1)$ e $d = Q_k(\alpha)/2\sqrt{n}$. Com essas modificações, obtiveram-se probabilidades de cobertura próximas ao valor nominal, para amostras pequenas e moderadas.

Para identificar quais proporções binomiais causaram a rejeição da hipótese de nulidade em relação a um valor de referência específico π_0 , Krishnamoorthy et al. (2004) propuseram um procedimento de construção de intervalos de confiança simultâneos para $\pi_i, i = 1, 2, \dots, k$. Os limites inferiores e superiores dos intervalos foram obtidos de uma distribuição beta com parâmetros (y_i, n_i) , tal que: $L_i = \text{Beta}(c/2, y_i, n_i - y_i + 1)$ e $U_i = \text{Beta}(1 - c/2, y_i + 1, n_i - y_i)$, mas, se $y_i = 0$, $(L_i, U_i) = (0, (1 - c)^{1/n_i})$ e se $y_i = n_i$, $(L_i, U_i) = (c^{1/n_i}, 1)$, sendo c o quantil da distribuição beta. Desse modo, os autores verificaram que, escolhendo $c = 1 - (1 - \alpha)^{1/k}$, o intervalo de confiança contém π_i com probabilidade de, pelo menos, $(1 - \alpha)^{1/k}$ e, então, $P(\pi_i \in (L_i, U_i), i = 1, 2, \dots, k) \geq 1 - \alpha$. Assim, se $\pi_0 \notin (L_j, U_j), j = 1, 2, \dots, k_l \leq k$, eles concluíram que os π_j 's são significativamente diferentes de π_0 , ao nível nominal α . Esse método foi ilustrado por meio da aplicação de dois exemplos.

McCann & Tebbs (2007) estenderam os procedimentos assintóticos de Hochberg & Tamhane (1987) e de Jeffreys-Perks, apresentados por Piegorsch (1991) para dados combinados. Nesse caso, as observações foram analisadas em grupo e não individualmente, de modo que, um grupo foi classificado como positivo se ao menos um indivíduo era positivo, considerando tamanhos de grupos maiores do que 1 ($s > 1$). Por exemplo, em estudos envolvendo análises de sangue são coletadas amostras de vários indivíduos. Essas amostras são combinadas e, em seguida, é realizado um único teste para verificar se o sangue desses indivíduos pode ser considerado positivo ou negativo, ao invés de realizar um teste para cada indivíduo.

Avaliando o intervalo de confiança simultâneo de Hochberg & Tamhane (1987) esses autores verificaram que esse intervalo apresentou excelentes probabilidades de cobertura para tamanhos de grupos maiores do que 10 ($s \geq 10$) e, observou-se que o intervalo aproximado de Jeffreys-Perks apresentou resultados conservativos quando tamanhos de grupos menores foram estabelecidos. Esses dois procedimentos de comparações múltiplas de pares de proporções estimadas, foram ilustrados utilizando dados de um estudo HIV observacional envolvendo homens que usam drogas intravenosas.

Agresti et al. (2008) apresentaram um método de construção de intervalos de confiança simultâneos, que utiliza a distribuição da amplitude estudentizada com uma estatística escore, para comparar proporções de várias amostras binomiais independentes. Esse método consiste em usar o intervalo de Wald, depois de adicionar um sucesso e um fracasso na amostra da população, e substituindo o percentil normal $Q_{k,\infty}(\alpha)/\sqrt{2}$ que, deve ser multiplicado pelo erro padrão, sendo $Q_{k,\infty}(\alpha)$ o quantil $100(1 - \alpha)\%$ da distribuição com um número infinito de graus de liberdade (amplitude padronizada). Esse método é aplicável a uma série de medidas como a razão de chances, diferenças entre proporções e o risco relativo.

Para a razão de chances, um estudo de simulação indicou que esse método tem probabilidade de cobertura mais próxima ao valor nominal do que o método de Bonfer-

roni aplicado ao intervalo de confiança padrão. Para a diferença de proporções binomiais $\pi_i - \pi_{i'}$, foi proposto o seguinte intervalo de confiança: $\tilde{\pi}_i - \tilde{\pi}_{i'} \pm \{Q_{k,\infty}(\alpha)/\sqrt{2}\} \sqrt{\frac{\tilde{\pi}_i(1-\tilde{\pi}_i)}{n_i+2} + \frac{\tilde{\pi}_{i'}(1-\tilde{\pi}_{i'})}{n_{i'}+2}}$, em que $\tilde{\pi}_i = \frac{y_i+1}{n_i+2}$ e $\tilde{\pi}_{i'} = \frac{y_{i'}+1}{n_{i'}+2}$. 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.

2.4.2 Teste de comparações múltiplas via *bootstrap*

Na literatura, são encontrados vários trabalhos envolvendo métodos de comparações múltiplas das médias de níveis dos fatores para populações normais e não normais, utilizando métodos de *bootstrap*. No entanto, o número de trabalhos envolvendo comparações múltiplas de proporções binomiais é relativamente pequeno.

Biase (2006) realizou comparações múltiplas em populações binomiais utilizando a técnica de *bootstrap* infinito, introduzida por Conlon & Thomas (1990) por meio de simulação Monte Carlo. Esses autores avaliaram as taxas de erro tipo I por experimento e o poder de dois testes de *bootstrap* infinito, um considerando o estimador de máxima verossimilhança (MV) e o outro o estimador de Pan (Pan). Esse último tem como característica a utilização de quatro pseudo-observações, sendo duas delas consideradas como sucesso do evento de interesse. Foram simuladas situações considerando tamanhos amostrais (n) iguais a 10, 30 e 100 e número de populações binomiais (k) iguais a 2, 5 e 10. Sob a hipótese de nulidade completa, admitiu-se que as probabilidades de sucesso (π) fossem iguais a 0,1, 0,5 e 0,9 e, sob a hipótese de nulidade parcial, considerou-se a formação de dois grupos, cuja diferença entre as probabilidades de sucesso dos grupos, denominada de Δ , variava entre 0,01 e 0,9.

A taxa de erro tipo I por experimento sob H_0 completa foi controlada em nível inferior ou, no máximo, igual ao valor nominal α adotado em ambos os testes. Assim,

em nenhuma configuração avaliada houve resultados que classificassem os testes como liberais. Para valores de π próximos a 0,5, houve tendências de melhores resultados dos testes, mesmo para tamanhos de amostras pequenas. Verificou-se um melhor desempenho do teste *bootstrap* MV devido à menor ocorrência de casos em que o teste foi considerado conservativo quando comparado com o teste *bootstrap* de Pan. Esses resultados são apresentados na Tabela 1.1, para o valor nominal de 5%. Para o valor nominal de significância de 1%, os resultados das taxas de erro tipo I por experimento foram bastante similares aos observados para 5%.

TABELA 1.1 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 os testes de *bootstrap* de Pan (Pan) e de máxima verossimilhança (MV), no valor nominal de 5%.

k	n	$\pi = 0,1$		$\pi = 0,5$		$\pi = 0,9$	
		Pan	MV	Pan	MV	Pan	MV
2	10	1,00 ⁺	1,00 ⁺	4,10 ^{ns}	4,05 ^{ns}	0,65 ⁺	0,65 ⁺
2	30	2,35 ⁺	3,95 ^{ns}	5,10 ^{ns}	4,90 ^{ns}	2,30 ⁺	3,95 ^{ns}
2	100	5,55 ^{ns}	5,85 ^{ns}	5,05 ^{ns}	4,90 ^{ns}	5,10 ^{ns}	5,20 ^{ns}
5	10	0,00 ⁺	0,15 ⁺	3,35 ⁺	3,30 ⁺	1,50 ⁺	1,00 ⁺
5	30	2,40 ⁺	3,90 ^{ns}	5,25 ^{ns}	5,15 ^{ns}	2,50 ⁺	4,20 ^{ns}
5	100	4,00 ^{ns}	4,55 ^{ns}	5,15 ^{ns}	5,25 ^{ns}	4,20 ^{ns}	4,30 ^{ns}
10	10	0,00 ⁺	1,35 ⁺	2,55 ⁺	2,55 ⁺	0,00 ⁺	0,80 ⁺
10	30	1,05 ⁺	3,15 ⁺	4,40 ^{ns}	4,20 ^{ns}	0,70 ⁺	3,00 ⁺
10	100	4,45 ^{ns}	5,20 ^{ns}	5,20 ^{ns}	5,20 ^{ns}	3,50 ⁺	4,35 ^{ns}

⁺ significativamente inferior, ao nível nominal de 5%, considerando uma confiança de 99%.

^{ns} não significativamente diferente do nível nominal de 5%, considerando uma confiança de 99%.

Quanto às taxas de erro tipo I por experimento sob H_0 parcial, observou-se que os dois testes de *bootstrap* apresentaram desempenhos similares em relação ao controle do erro tipo I e foram classificados, na grande maioria dos casos, como conservativos, tanto para $\alpha = 5\%$ quanto para $\alpha = 1\%$.

O poder do teste *bootstrap* MV apresentou pequena superioridade em relação ao poder do teste *bootstrap* de Pan, embora nas situações em que o teste *bootstrap* de Pan

foi superior, esta superioridade foi muito expressiva. Isso ocorreu quando os valores de Δ eram grandes ($\Delta \geq 0,8$) e os tamanhos amostrais eram menores ($n \leq 30$) sob a hipótese H_1 ($\pi_1 \neq \pi_2 \neq \dots \neq \pi_k$) e H_0 parcial ($\pi_1 = \pi_2 = \dots = \pi_i \neq \pi_{i+1} = \pi_{i+2} = \dots = \pi_k$). Os valores de poder de ambos os testes foram relativamente pequenos para pequenas amostras ($n \leq 30$) e, com o aumento do número de populações, constatou-se uma redução expressiva dos valores de poder de ambos os testes. Essa redução tornou-se menor à medida que o tamanho das amostras aumentava. Esses fatos ocorreram tanto sob H_1 quanto sob H_0 parcial.

Westfall & Young (1989), realizando estudos clínicos no desenvolvimento de novas drogas, modeladas por meio da distribuição binomial multivariada, propuseram o uso de valores-p ajustados nos PCM, utilizando métodos de reamostragem *bootstrap*. Esta metodologia foi comparada ao ajustamento de Bonferroni usual e os autores certificaram que, com esse ajustamento, os valores-p são excessivamente conservativos para muitas situações envolvendo testes com dados binomiais. Os autores verificaram que os resultados do ajustamento de reamostragem de *bootstrap* e permutação foram similares, principalmente para grandes amostras.

Jhun & Jeong (2000) construíram regiões de confiança simultâneas para proporções de uma única população multinomial e para um número finito de contrastes de várias populações multinomiais utilizando métodos de *bootstrap*. O desempenho dos métodos de *bootstrap* foi comparado ao dos métodos de Goodman (1964) e Bonferroni, em relação à probabilidade de cobertura média via simulação Monte Carlo. Esses autores concluíram que os métodos de *bootstrap* propostos apresentaram pequenas vantagens em relação aos outros, pois são mais precisos, em termos de probabilidade de cobertura média.

2.4.3 Teste de comparações múltiplas bayesianos

Para solucionar os problemas envolvidos na aplicação dos PCM, métodos alternativos de comparações múltiplas são encontrados na literatura. Esses métodos envolvem metodologias bayesianas.

O primeiro método de comparação múltipla bayesiano foi proposto por Duncan (1965). Em uma primeira etapa da aproximação bayesiana, o método de Duncan consiste em especificar uma distribuição *a priori* conjunta para os parâmetros e assumir um modelo usual de efeitos aleatórios para as médias θ_i , estabelecendo uma relação conhecida para a razão da variância entre e dentro, $k = \sigma_\theta^2/\sigma^2$, obtendo-se, assim, a distribuição *a posteriori* das médias. Na segunda etapa, uma das mais importantes contribuições de Duncan (1965) foi estabelecida. Ele modelou a consequência de duas ou mais médias serem iguais usando funções de perda e verificou que, ao assumir a mesma função perda entre pares de comparações, era necessário especificar apenas uma constante k e isso revela uma importância relativa do erro tipo I em relação ao erro tipo II, para cada par de comparação.

Waller & Duncan (1969) modificaram o método de Duncan original usando uma *priori* hierárquica para σ_θ^2/σ^2 . A vantagem dessa alteração é que não é necessário especificar um valor para a razão desconhecida de variâncias entre e dentro. Esses autores estabeleceram uma conexão entre os possíveis valores de k e o nível de significância envolvido na comparação.

A partir da década de 1990, muitos estudos envolvendo PCM têm aplicado a metodologia bayesiana com sucesso. Alguns deles combinam idéias bayesianas e frequentistas.

Consonni & Veronese (1995) consideraram um conjunto de experimentos binomiais no intuito de inferir sobre um ou mais π_i , em que π_i é a probabilidade de sucesso correspondente ao experimento i . Os autores utilizaram distribuições *a priori* para π_i mais flexíveis, considerando vários graus de similaridade entre os π_i , seme-

lhante a técnica de cluster. Por meio de conjunto de dados reais, eles verificaram que essa metodologia pode ser aplicada favoravelmente às análises de variáveis respostas binárias na presença de covariáveis categóricas.

Gopalan & Berry (1998) afirmaram que os PCM estão entre os problemas mais difíceis encontrados por estatísticos e outros pesquisadores e que muitos desses PCM controlam apenas uma das taxas de erro tipo I: por comparação ou por experimento. Utilizando uma abordagem bayesiana ao problema de comparações múltiplas, esses autores utilizaram famílias do processo de Dirichlet como distribuições *a priori* para obter probabilidades *a posteriori* de várias hipóteses de igualdade entre médias populacionais, sob as duas diferentes combinações de *priori*/verossimilhança: beta/binomial e normal/gama invertida com variância iguais. Como as distribuições *a posteriori* eram complexas de serem executadas analiticamente, os autores utilizaram o algoritmo de simulação aproximado, conhecido por amostrador de Gibbs. O procedimento proposto foi comparado com o teste de Duncan, que mostrou ser mais poderoso sob certas hipóteses alternativas.

Berry & Hochberg (1999) mostraram que as inferências bayesiana e frequentista podem apresentar resultados similares para o problema de comparações múltiplas, quando a taxa de erro por experimento é controlada. Esses autores descreveram a diferença entre distribuição *a priori* independente e hierárquica e, como ilustração, utilizaram uma distribuição *a priori* de Dirichlet no contexto de ambiguidade dos resultados. E, finalmente, discutiram alguns procedimentos quasi-bayesiano que combinam ideias bayesiana e frequentista, mostrando que a metodologia bayesiana tem potencial para determinar procedimentos que podem ser avaliados utilizando-se critérios objetivos.

Bratcher & Hamilton (2005) propuseram um procedimento bayesiano de comparações múltiplas para avaliar médias de populações normais com variâncias homogêneas, considerando um modelo de função perda constante. Por meio de simulações, compararam o desempenho desse modelo com o bayesiano, usando função perda linear

e com os métodos frequentistas usuais e verificaram que, para o caso de *prioris* não informativas, o método proposto apresentou melhor desempenho do que os demais, além de ser apropriado para todos os tamanhos amostrais.

Scott & Berger (2006), motivados pela necessidade de analisar dados de DNA, exploraram vários aspectos dos PCM utilizando a abordagem bayesiana, apresentando alguns exemplos de aplicação com esse tipo de dados.

Ali et al. (2006) estudaram as comparações múltiplas bayesianas em populações binomiais negativas utilizando *prioris* da família de Dirichlet, o que possibilitou obter probabilidades *a posteriori* para diferentes hipóteses sobre os parâmetros da distribuição binomial negativa. O cálculo das probabilidades *a posteriori* foi realizado utilizando-se o amostrador de Gibbs devido à dificuldade de se encontrar formas analíticas.

Andrade (2008) propôs a utilização de procedimentos bayesianos para realizar comparações múltiplas em populações normais homocedásticas e heterocedásticas via simulação Monte Carlo. A partir da distribuição *a posteriori*, foram geradas, sob H_0 , k cadeias de médias, assumindo médias constantes, e obtida a amplitude padronizada da *posteriori*, por meio da distribuição *a posteriori* das médias. Para a realização das inferências, sob H_0 , obtiveram-se a diferença mínima significativa e o intervalo de credibilidade bayesiano. Concluiu-se, então, que os PCM bayesianos foram propostos com sucesso, pois os procedimentos baseados na amplitude padronizada foram superiores aos demais procedimentos estudados, por terem controlado o erro tipo I e detectado a maior parte das diferenças sob H_1 , nos exemplos simulados.

2.5 Inferência bayesiana

Durante grande período, a metodologia bayesiana ficou resguardada por necessitar de resoluções matemáticas inviáveis de serem obtidas analiticamente, mais especificamente de integrações. Por volta da década de 1960, a análise bayesiana ressurgiu

em alguns trabalhos teóricos, como o de Jeffreys (1961), mas somente em 1990, Gelfand & Smith (1990) conseguiram solucionar o problema das integrações de maneira alternativa, utilizando um recurso de simulação dinâmica, denominado algoritmo *Gibbs Sampler*.

O avanço dos recursos computacionais possibilitou que a abordagem bayesiana fosse aplicada com maior intensidade, pois, com esses recursos, foi possível implementar técnicas de simulação intensiva que favoreceram a solução aproximada de problemas que, anteriormente, eram de difícil solução.

Atualmente, a inferência bayesiana é uma das duas possíveis alternativas aos procedimentos clássicos de estimação e testes de hipóteses.

Na inferência bayesiana, o parâmetro θ é desconhecido e toda incerteza a seu respeito deve ser quantificada em termos de probabilidade, sendo de interesse tentar reduzir essa incerteza por meio das informações obtidas de experiências anteriores ou do conhecimento do pesquisador na área em questão (Gelman et al., 1997). No modelo bayesiano, o parâmetro θ é também considerado uma variável aleatória, ao contrário do modelo clássico, que considera o parâmetro como um valor fixo ou constante, ignorando toda a informação do pesquisador. A inferência clássica assume que todas as informações ou inferências sobre os parâmetros de interesse sejam obtidas a partir de dados amostrais selecionados aleatoriamente da população.

Segundo Box & Tiao (1992), a inferência bayesiana considera toda a informação do pesquisador sobre o parâmetro de interesse θ , fundamentado em algum conhecimento *a priori*. Essa informação é representada por uma função de distribuição $p(\theta)$, conhecida por distribuição *a priori*.

Os dados $\mathbf{y} = \{ y_1, y_2, \dots, y_n \}$, representados por uma amostra aleatória de uma população com densidade f , são considerados na análise bayesiana, por meio da função de verossimilhança. Essa função é denotada por $L(\mathbf{y}|\theta)$ e constitui a densidade conjunta dos dados (Paulino et al., 2003).

Assim, a inferência bayesiana consiste do conhecimento prévio em relação aos parâmetros (distribuição *a priori*) e das informações referentes aos dados amostrais (função de verossimilhança). A partir dessas informações, obtém-se a densidade *a posteriori* dos parâmetros a serem estimados. O mecanismo utilizado para combinar a distribuição *a priori* e a função de verossimilhança é o Teorema de Bayes, que permite obter a distribuição de densidade *a posteriori*, $p(\theta|\mathbf{y})$, dada por:

$$p(\theta|\mathbf{y}) = \frac{L(\mathbf{y}|\theta)p(\theta)}{\int L(\mathbf{y}|\theta)p(\theta)d\theta}. \quad (2.2)$$

Na expressão (2.2), o denominador não depende de θ e funciona como uma constante normalizadora de $p(\theta|\mathbf{y})$, pois depende somente da amostra dada. Portanto, a expressão do Teorema de Bayes pode ser simplificada por:

$$p(\theta|\mathbf{y}) \propto L(\mathbf{y}|\theta)p(\theta),$$

em que \propto representa proporcionalidade.

Desse modo, a densidade *a posteriori* de θ incorpora, via Teorema de Bayes, toda a informação disponível sobre o parâmetro e é proporcional ao produto da função de verossimilhança e a densidade *a priori* de θ (Gelman et al., 1997).

De acordo com Broemiling (1989), o Teorema de Bayes pode ser visto como um método de atualização da opinião do pesquisador sobre o parâmetro θ e pode ser considerado a base da inferência bayesiana, pois todas as inferências a respeito dos parâmetros é realizada a partir da distribuição *a posteriori* obtida.

A função de verossimilhança é muito importante no Teorema de Bayes, pois, por meio dela, os dados modificam o conhecimento que se tem *a priori* sobre θ e pode, entretanto, ser considerada como a representação da informação de θ obtida por meio dos dados (Box & Tiao, 1992).

2.5.1 Distribuição *a priori*

A distribuição *a priori* é a distribuição de probabilidade que consegue captar a variabilidade dos graus de conhecimento de uma quantidade de interesse θ . Essa distribuição é especificada por meio de experiências ou crença dos pesquisadores e é considerada de natureza subjetiva, ou seja, varia de problema para problema e de pesquisador para pesquisador (Paulino et al., 2003).

As distribuições *a priori* podem ser informativas ou não informativas e são descritas a seguir.

As *prioris* informativas são utilizadas quando o pesquisador tem algum conhecimento prévio sobre o parâmetro em questão, que são incorporadas na análise por meio da distribuição *a priori* $p(\theta)$ que mais bem o representa. Essa distribuição *a priori*, em algumas situações, é especificada com o auxílio de constantes, denominadas de hiperparâmetros, que representam os parâmetros da distribuição dos parâmetros de interesse θ (Paulino et al., 2003).

Uma classe de *prioris* interessantes ocorre quando as distribuições *a priori* e *a posteriori* pertencem à mesma classe de distribuições. Quando isso acontece, diz-se que as *prioris* são conjugadas. Nesse caso, a atualização do conhecimento sobre o parâmetro θ pode ser realizada alterando-se somente os valores dos hiperparâmetros (Gelman et al., 1997). Esses autores comentam, ainda, que a família de distribuições conjugadas é computacionalmente conveniente, pelo fato de apresentar uma forma paramétrica conhecida. Como exemplificação, eles mostraram que a família da distribuição beta é conjugada ao modelo bernoulli e binomial.

As *prioris* não informativas são utilizadas quando o pesquisador tem pouco ou nenhum conhecimento sobre o parâmetro. Em outras palavras, as informações *a priori* são relativamente não significativas em relação às informações amostrais (Paulino et al., 2003).

Segundo Box & Tiao (1992), quando se utilizam *prioris* não informativas, pode-

se supor que todos os possíveis valores do parâmetro são igualmente prováveis, ou seja, com uma distribuição uniforme. Nesse caso, a distribuição *a priori* será proporcional a uma constante ($p(\theta) \propto k$), o que equivale a dizer que nenhum valor particular de θ tem preferência. Outro conjunto de *priori* não informativas utilizadas com bastante frequência na literatura refere-se à classe de *prioris* de Jeffreys (Gelman et al., 1997).

De acordo com Paulino et al. (2003), a princípio, *as prioris* não informativas eram interpretadas como representações formais de ignorância, mas, atualmente, essas *prioris* são consideradas opções convencionais de defeito a que se recorre em caso de informação *a priori* insuficiente, que torna difícil nomear uma distribuição subjetiva considerada adequada.

2.5.2 Métodos de simulação Monte Carlo via Cadeias de Markov (MCMC)

Quando se pretende realizar a inferência sobre a densidade conjunta *a posteriori*, o interesse consiste em encontrar uma distribuição para um parâmetro específico θ_i , $i = 1, 2, \dots, k$. Essa distribuição é conhecida como distribuição marginal e sua obtenção baseia-se na integração da densidade *a posteriori* em relação aos outros parâmetros do modelo, ou seja,

$$p(\theta_i|\mathbf{y}) = \int p(\theta_1, \theta_2, \dots, \theta_k|\mathbf{y}) d\theta_{-i}$$

em que: $\theta_{-i} = (\theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_k)$ é o vetor de parâmetros com o i -ésimo componente removido.

A distribuição marginal de um parâmetro θ contém toda a informação probabilística a seu respeito. No entanto, a resolução dessa integral é, na maioria das vezes, complexa ou, até mesmo, impossível de ser executada analiticamente. Portanto, a inferência exata somente será possível se essas integrais puderem ser calculadas analiticamente. Caso contrário, uma das alternativas existentes são os métodos aproximados de

inferência, conhecidos como métodos de simulação Monte Carlo via Cadeias de Markov (MCMC). Para a utilização desses métodos, é necessário que se obtenha, da distribuição *a posteriori*, um conjunto de distribuições, denominadas distribuições condicionais completas (Paulino et al., 2003).

A distribuição condicional completa do parâmetro θ_i é obtida considerando que, na densidade conjunta, os demais parâmetros θ_{-i} são conhecidos. Desta forma, as constantes podem ser desprezadas e a expressão se torna menos complexa.

Os métodos de MCMC são processos estocásticos que consideram as distribuições condicionais completas *a posteriori* de cada parâmetro θ_i , $i = 1, 2, \dots, k$, para gerar amostras que convergem para a densidade marginal à medida que o tamanho da amostra aumenta, por meio de simulação Monte Carlo (Gelfand, 2000). O Amostrador de Gibbs e o Metropolis-Hastings são algoritmos que utilizam esses métodos e serão especificados a seguir.

O Amostrador de Gibbs foi introduzido por Gelman & Gelman (1984), para simular distribuições multivariadas de natureza bastante complexa e constitui uma ferramenta de grande importância na resolução de problemas em inferência bayesiana (Paulino et al., 2003).

O algoritmo do Amostrador de Gibbs é, essencialmente, um método iterativo de amostragem de uma cadeia de Markov, cujo núcleo de transição é constituído pelas distribuições condicionais completas. Portanto, esse algoritmo é empregado quando as distribuições condicionais completas apresentam formas de densidades conhecidas que, nesse caso, são facilmente amostradas, por meio da geração de variáveis aleatórias que convergem para a distribuição marginal, mesmo quando a sua densidade não é conhecida (Gelman, 1997).

Se o Amostrador de Gibbs não apresenta resultados eficientes, situação verificada quando a distribuição condicional completa dos parâmetros não apresenta uma expressão de densidade conhecida, um algoritmo denominado por Metropolis-Hastings

pode ser utilizado, para obter a distribuição marginal *a posteriori*. Nesse caso, os valores do parâmetro são gerados de uma distribuição proposta e esses valores serão aceitos ou não com uma certa probabilidade de aceitação (Chib & Greenberg, 1995).

Para a aplicação dos algoritmos Amostrador de Gibbs e Metropolis-Hastings, algumas considerações devem ser observadas, tais como o diagnóstico da convergência, o período de descarte amostral (*burn-in*), o intervalo de amostragem (*thin*) e o tamanho da cadeia amostral.

Na prática, admite-se a convergência quando a série alcança um estado de estacionariedade, o que significa que as distribuições condicionais completas estão suficientemente próximas das distribuições marginais. Como os algoritmos MCMC são processos iterativos, o grande problema está em encontrar o número de iterações necessárias para que essa convergência seja verificada (Gamerman, 1997). Na presença de modelos complicados, os algoritmos MCMC necessitam de um grande esforço computacional. O ideal é obter a convergência fazendo o menor esforço computacional possível e, para que isso ocorra, existem alguns métodos formais na literatura para avaliar essa convergência. Dentre os mais relevantes, destacam-se os de Geweke (1992), Gelman & Rubin (1992), Heidelbberg & Welch (1983) e Raftery & Lewis (1992).

2.5.3 Estimação usando a inferência bayesiana

Na teoria frequentista, podem realizar inferências por meio dos processos de estimação, pontual e intervalar, e dos testes de hipóteses. Esses mesmos procedimentos também podem ser utilizados para realizar inferências na abordagem bayesiana. Do ponto de vista bayesiano, a distribuição *a posteriori* de um parâmetro dado os dados fornece inferências completas (Bolstad, 2004).

A distribuição marginal *a posteriori* de um parâmetro θ contém toda a informação probabilística de θ . Algumas vezes, no entanto, é necessário resumir a informação contida nessa distribuição por meio de alguns valores numéricos. O caso mais simples

é a estimação pontual de θ , que resume a distribuição marginal *a posteriori* por meio de um único valor de θ . Esse valor é denotado por $\hat{\theta}$ e denominado de estimador pontual de θ (Box & Tiao, 1992).

A escolha das estimativas bayesianas de θ depende naturalmente da expressão de $p(\theta|\mathbf{y})$ e dos objetivos da pesquisa. Dentre as estimativas pontuais mais usuais, Gelman et al. (1997) mencionam a média, a moda e a mediana da distribuição *a posteriori*. A média é interpretada como a esperança *a posteriori* do parâmetro θ . A moda *a posteriori*, de acordo com Paulino et al. (2003), coincide com a estimativa de máxima verossimilhança de θ , quando a distribuição *a priori* é constante e pode ser interpretada como o valor de θ que tem a maior credibilidade *a posteriori* e não como o valor de θ que torna mais plausível a amostra observada, como interpretada na inferência clássica.

A estimação pontual é mais bem caracterizada na teoria da decisão, em que é muito comum, na inferência bayesiana, aplicar funções de perda ou de risco para gerar critérios de decisão. Essas funções, no caso de testes de hipóteses, atribuem perdas conforme a decisão tomada, dado que uma determinada hipótese é verdadeira. O risco ou a perda esperada *a posteriori* são obtidos tomando-se a esperança da função perda em relação à distribuição *a posteriori* de θ . Uma regra de decisão é considerada ótima quando se obtém um risco mínimo (Box & Tiao, 1992).

Segundo Gelman et al. (1997), a principal restrição da estimação pontual é que, ao estimar um parâmetro por meio de um único valor numérico, toda a informação presente na distribuição *a posteriori* é resumida por esse número. Esse tipo de estimação não especifica a magnitude do erro cometido e, portanto, não é possível ter uma idéia da precisão associada ao estimador.

Uma informação da distribuição *a posteriori* mais precisa do que qualquer estimativa pontual é obtida de uma região do espaço paramétrico Θ que contenha uma parte substancial da massa probabilística *a posteriori* (Paulino et al., 2003). Assim, para obter essa informação, pode-se realizar a inferência utilizando-se os intervalos de

credibilidade, que são intervalos alternativos aos intervalos de confiança clássico.

Gelman et al. (1997) definem C como um intervalo de credibilidade $100(1 - \alpha)\%$ para θ se $P(\theta \in C) \geq (1 - \alpha)$. Quanto menor for o tamanho desse intervalo, mais concentrada é a distribuição do parâmetro θ , ou seja, o tamanho do intervalo informa sobre a dispersão do parâmetro.

Uma infinidade de intervalos de confiança pode ser obtida. No entanto, o interesse principal está naquele com o menor comprimento possível. Os intervalos de comprimento mínimo são obtidos tomando-se os valores de θ com a maior densidade *a posteriori* e são denominados de intervalos de credibilidade de máxima densidade *a posteriori* (HPD - *Highest posterior density interval*).

Um intervalo de credibilidade C de $100(1 - \alpha)\%$ para θ é de máxima densidade *a posteriori* se $C = \{\theta \in \Theta : p(\theta|y) \geq k(\alpha)\}$, em que $k(\alpha)$ é a maior constante obtida, tal que $P(\theta \in C) \geq 1 - \alpha$ (Gelman et al., 1997; Paulino et al., 2003; Box & Tiao, 1992).

Os intervalos de credibilidade e o HPD são coincidentes, quando as distribuições *a posteriori* são unimodais e simétricas. No caso de distribuições assimétricas, esses intervalos apresentam diferenças, que dependem do grau de assimetria. Finalmente, quando a distribuição *a posteriori* é multimodal, diversos subintervalos em torno das modas mais relevantes podem ser obtidos e, nessas condições, o intervalo HPD apresenta certa vantagem em relação ao intervalo de credibilidade pelo fato de fornecer maiores informações (Gelman et al., 1997).

2.6 Teste de razão de verossimilhanças

O teste de razão de verossimilhanças é um procedimento completamente geral de obtenção da estatística teste em qualquer situação, univariada ou multivariada, para o qual é possível maximizar a verossimilhança das observações. Este teste é utilizado

quando se deseja testar a hipótese de nulidade H_0 contra a hipótese global alternativa H_1 (Mardia et al., 1995).

Suponha, então, que o interesse esteja em testar a hipótese H_0 de que um parâmetro θ pertença a algum subespaço de \mathbb{R}^s . Este subespaço é conhecido como conjunto nulo e é representado por $\Omega_0 \subset \mathbb{R}^s$. Geralmente, este subespaço corresponde às restrições que são impostas no espaço paramétrico e, portanto, neste caso, a hipótese nula equivale ao espaço restrito. A solução deste teste de hipótese, em termos da região de rejeição R , é um conjunto de valores do espaço amostral que levam à decisão de rejeitar a hipótese H_0 em favor da hipótese alternativa H_1 , que é denominada de espaço irrestrito (Ferreira, 2008).

A região de rejeição R do teste de razão de verossimilhanças baseia-se no controle do erro tipo I sob um valor pré-fixado de $100\alpha\%$, ou seja, $P(\text{rejeitar } H_0 | H_0 \text{ verdadeira}) = \alpha$. Desse modo, R é determinada por:

$$\sup_{\theta \in \Omega_0} P(\mathbf{y} \in R; \theta) = \alpha,$$

em que \mathbf{y} é o vetor $1 \times k$ de dados e α é o nível nominal de significância do teste ($0 < \alpha < 1$) (Mardia et al., 1995; Ferreira, 2008).

Assim, se a distribuição da amostra aleatória \mathbf{y} depende de um parâmetro θ , e se $H_0 : \theta \in \Omega_0$ e $H_1 : \theta \in \Omega$ são quaisquer duas hipóteses, em que H_0 pertence ao espaço restrito Ω_0 e H_1 ao espaço irrestrito Ω , então, a estatística da razão de verossimilhanças, para testar H_0 contra H_1 , é definida por Mardia et al. (1995), por:

$$\Lambda = \frac{L_{\Omega_0}(\mathbf{y}; \hat{\theta})}{L_{\Omega}(\mathbf{y}; \hat{\theta})}$$

em que $L_{\Omega_0}(\mathbf{y}; \hat{\theta})$ é o máximo da função de verossimilhança para o espaço restrito e $L_{\Omega}(\mathbf{y}; \hat{\theta})$ é o máximo da função de verossimilhança para o espaço irrestrito.

Quando o valor da razão de verossimilhanças em seu máximo é grande, é mais

provável que a hipótese H_0 não será rejeitada. Caso contrário, se a razão de verossimilhanças for pequena, a hipótese H_1 deve ser escolhida (Johnson & Wichern, 2002; Ferreira, 2008).

De acordo com Ferreira (2008), pode-se estabelecer uma região de rejeição de H_0 com base na distribuição de Λ , para especificar um teste de tamanho α . No entanto, obter a distribuição nula de Λ é tarefa bastante complicada.

Neste contexto, se $\Omega_0 \subset \Omega$, com $\Omega_0 \subset \mathbb{R}^s$ e $\Omega \subset \mathbb{R}^r$ e, sob determinadas condições de regularidade, para cada $\boldsymbol{\theta} \in \Omega_0$ $-2\ln(\Lambda)$ tem distribuição assintoticamente de qui-quadrado com $r - s$ graus de liberdade (Mood et al., 1974; Mardia et al., 1995). Portanto, a região de rejeição da hipótese nula para o teste da razão de verossimilhanças é dada por:

$$R = \{\mathbf{y} | \lambda = -2\ln[\Lambda(\mathbf{y})] > \chi_{\alpha, r-s}^2\}$$

em que $\chi_{\alpha, r-s}^2$ é o quantil superior da distribuição qui-quadrado com $r - s$ graus de liberdade.

2.6.1 Estimação

A inferência estatística é realizada por meio dos processos de estimação e dos testes de hipóteses e tem por objetivo obter informações dos valores dos parâmetros desconhecidos em relação às amostras selecionadas da população (Walpole, 1974).

Os processos de estimação permitem obter aproximações numéricas para os parâmetros. Segundo Mood et al. (1974), existem, basicamente, dois processos de estimação: a estimação pontual e a estimação intervalar.

A estimação pontual ou por ponto é aquela em que se obtém, por meio de um estimador, um único valor amostral para estimar o parâmetro populacional. Convém ressaltar que o estimador é uma variável aleatória, que é função dos elementos amostrais. Portanto, a estimativa pontual pode variar entre as amostras, pois está sujeita a

erros de estimação, devido ao processo de aleatorização e à variabilidade inerente à população da qual a amostra foi selecionada.

Desse modo, a estimação pontual não fornece idéia da margem de erro que é cometida ao se estimar um determinado parâmetro. Por essa razão, é importante obter a estimação por intervalo, que procura suprir essa necessidade (Ferreira, 2005).

A estimação intervalar consiste em construir um intervalo com uma probabilidade pré-fixada de conter o verdadeiro valor do parâmetro desconhecido. Geralmente, essas probabilidades são fixadas em 95% ou 99% e são conhecidas por coeficiente de confiança (Mood et al., 1974).

Algumas propriedades dos estimadores são desejáveis na teoria da inferência. A escolha de um estimador de um parâmetro θ qualquer em relação a outro depende de uma avaliação criteriosa dessas propriedades (Ferreira, 2005). Essas principais propriedades são citadas a seguir:

- um estimador $\hat{\theta}$ é considerado não viesado quando sua esperança matemática é igual ao valor do parâmetro populacional θ , isto é, $E(\hat{\theta}) = \theta$;
- o estimador de maior eficiência, dentre todos os estimadores não viesados de θ , é aquele que possui menor variância;
- um estimador é consistente se, além de ser não viesado, sua variância tende a zero, quando o tamanho da amostra n aumenta, ou seja, $\lim_{n \rightarrow \infty} \sigma_{\hat{\theta}}^2 = 0$.

O método da máxima verossimilhança é um dos métodos de estimação pontual mais importantes e utilizados na teoria estatística (Mood et al., 1974). Neste método, os estimadores são obtidos a partir da maximização da função de verossimilhança (Bolfarine & Sandoval, 2000).

Para ilustrar o procedimento de obtenção do estimador de máxima verossimilhança, considere uma amostra aleatória Y_1, Y_2, \dots, Y_n de uma população com densidade $f(y)$, dependente do parâmetro θ . Em razão de os valores amostrais Y_1, Y_2, \dots ,

Y_n serem independentes, é possível definir a densidade conjunta ou função de verossimilhança (L) pelo produtório das densidades de cada Y_i ($i = 1, 2, \dots, n$). Define-se, então, a função de verossimilhança:

$$L = f(y_1) \cdot f(y_2) \cdot f(y_3) \cdot \dots \cdot f(y_n) = \prod_{i=1}^n f(y_i)$$

O estimador de máxima verossimilhança ($\hat{\theta}$) é aquele que maximiza o valor de L (Ferreira, 2005).

Após obter a expressão da função de verossimilhança, o estimador de máxima verossimilhança $\hat{\theta}$ é obtido tomando-se a primeira derivada de L em relação ao parâmetro θ , igualando a zero e resolvendo a expressão para θ (Mood et al., 1974).

Nas situações que a função de verossimilhança contém mais de um parâmetro, devem-se tomar as derivadas parciais de L em relação a cada um deles. Dessa forma, iguala-se cada derivada a zero e resolve-se o sistema formado, obtendo-se, assim, os estimadores de máxima verossimilhança dos parâmetros.

Existem ainda algumas propriedades matemáticas do logaritmo da função de verossimilhança que garantem a possibilidade de usar uma função suporte $S = \ln(L)$ em seu lugar, uma vez que apresentam o máximo para o mesmo valor de θ . Isso facilita a obtenção do máximo, pois, ao aplicar o logaritmo neperiano, o produtório transforma-se em somatório (Ferreira, 2005).

No caso específico deste trabalho, o interesse está voltado aos estudos de proporções binomiais independentes, cujos parâmetros são as probabilidades de sucesso (π_i) e o tamanho das amostras (n_i), $i = 1, 2, \dots, k$. No entanto, apenas os parâmetros π_i são desconhecidos e precisam ser estimados. A descrição de como obter os estimadores de máxima verossimilhança para esses parâmetros está apresentada na seção 4.1.

2.7 Inferências sobre proporções multinomiais

Conforme mencionado anteriormente, o foco deste trabalho está no desenvolvimento de metodologias envolvendo populações binomiais independentes. Entretanto, serão apresentados a seguir procedimentos de inferência para modelos multinomiais, pois, para modelos binomiais independentes, o raciocínio é análogo, uma vez que os modelos multinomiais constituem uma generalização dos modelos binomiais.

Considere, então, uma amostra aleatória de tamanho n obtida de uma população classificada em k categorias mutuamente exclusivas. Cada unidade amostral independente e identicamente distribuída pode ter realizações em qualquer das k categorias. Seja X_{ij} a variável correspondente à j -ésima unidade amostral da i -ésima categoria, cujas possíveis realizações são 1 para o sucesso e 0 para o fracasso. Então, $\mathbf{X}_j^\top = [X_{1j}, X_{2j}, \dots, X_{kj}]$ é uma observação multinomial com $\sum_{i=1}^k X_{ij} = 1$ e tem-se também que X_{ij} é linearmente dependente das demais $k - 1$ categorias. Se a amostra aleatória de tamanho n desta população for obtida utilizando-se n independentes ensaios multinomiais, tem-se que $Y_i = \sum_{j=1}^n X_{ij}$, cuja realização é representada por n_i , $i = 1, 2, \dots, k$. As contagens $\mathbf{Y} = [Y_1, Y_2, \dots, Y_k]^\top$ possuem distribuição multinomial.

Desse modo, seja $\pi_i = P(X_{ij} = 1)$ a probabilidade de sucesso na i -ésima categoria em uma realização j qualquer do experimento. Segundo Ferreira (2008), a função de probabilidade multinomial é dada por:

$$P(Y_1 = n_1, Y_2 = n_2, \dots, Y_k = n_k) = \frac{n!}{n_1!n_2! \dots n_k!} \prod_{i=1}^k \pi_i^{n_i}, \quad (2.3)$$

sendo que $\sum_{i=1}^k n_i = n$ e $\sum_{i=1}^k \pi_i = 1$.

Nessas condições, a média e a variância das variáveis Y_i são dadas, respectivamente, por $n\pi_i$ e $n\pi_i(1 - \pi_i)$ e a covariância entre as variáveis Y_i e $Y_{i'}$ é igual a $-n\pi_i\pi_{i'}$. Para inferir sobre os parâmetros desconhecidos π_i , obteve-se, por meio da

expressão (2.3), o estimador de máxima verossimilhança:

$$\hat{\pi}_i = \frac{Y_i}{n}.$$

Utilizando-se esse estimador, é possível obter o valor esperado, a variância e a covariância do estimador de π . Observa-se que $\hat{\pi}_i$ é obtido dividindo a variável aleatória por uma constante n . Então, seu valor esperado e a sua covariância são dados, respectivamente, por:

$$E(\hat{\boldsymbol{\pi}}) = \begin{bmatrix} \pi_1 \\ \pi_2 \\ \vdots \\ \pi_k \end{bmatrix} \quad \text{e} \quad \text{cov}(\hat{\boldsymbol{\pi}}) = \begin{bmatrix} \frac{\pi_1(1-\pi_1)}{n} & -\frac{\pi_1\pi_2}{n} & \cdots & -\frac{\pi_1\pi_k}{n} \\ -\frac{\pi_2\pi_1}{n} & \frac{\pi_2(1-\pi_2)}{n} & \cdots & -\frac{\pi_2\pi_k}{n} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{\pi_k\pi_1}{n} & -\frac{\pi_k\pi_2}{n} & \cdots & \frac{\pi_k(1-\pi_k)}{n} \end{bmatrix}. \quad (2.4)$$

Pelo Teorema do Limite Central, Ferreira (2008) considera, assintoticamente, que $\hat{\boldsymbol{\pi}}$ possui distribuição aproximadamente normal multivariada, ou seja, $\hat{\boldsymbol{\pi}} \sim N_p(\boldsymbol{\pi}, \text{cov}(\hat{\boldsymbol{\pi}}))$. O autor ressalta, ainda, que as matrizes $\text{cov}(\hat{\boldsymbol{\pi}})$ ou $\hat{\text{cov}}(\hat{\boldsymbol{\pi}})$ são singulares, pois $\sum_{i=1}^k \pi_i = 1$ e $\sum_{i=1}^k \hat{\pi}_i = 1$. Com isso, a distribuição multivariada degenera-se em uma dimensão inferior, $k - 1$. A matriz $\hat{\text{cov}}(\hat{\boldsymbol{\pi}})$ é obtida substituindo-se π por $\hat{\pi}$ na matriz da covariância em (2.4). Associando-se estas ideias com a teoria das formas quadráticas, pode-se obter a seguinte estatística, quando pretende-se testar a hipótese $H_0 : \boldsymbol{\pi} = \boldsymbol{\pi}_0$, sob a restrição de que $\boldsymbol{\pi}_0^\top \mathbf{1} = 1$, em que $\mathbf{1}$ é um vetor de uns:

$$\begin{aligned} \chi_c^2 &= n(\hat{\boldsymbol{\pi}} - \boldsymbol{\pi}_0)^\top \boldsymbol{\Sigma}_0^- (\hat{\boldsymbol{\pi}} - \boldsymbol{\pi}_0) \\ &= \sum_{i=1}^k \frac{(n_i - n\pi_{0i})^2}{n\pi_{0i}}, \end{aligned}$$

que, sob H_0 , possui distribuição assintoticamente qui-quadrado com $\nu = k - 1$ graus

de liberdade, sendo Σ_0^- definida por:

$$\Sigma_0^- = \begin{bmatrix} \frac{1}{\pi_{01}} + \frac{1}{\pi_{0k}} & \frac{1}{\pi_{0k}} & \dots & \frac{1}{\pi_{0k}} & 0 \\ \frac{1}{\pi_{0k}} & \frac{1}{\pi_{02}} + \frac{1}{\pi_{0k}} & \dots & \frac{1}{\pi_{0k}} & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{1}{\pi_{0k}} & \frac{1}{\pi_{0k}} & \dots & \frac{1}{\pi_{0k-1}} + \frac{1}{\pi_{0k}} & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Essa estatística de χ_c^2 é conhecida por X^2 de Pearson. Outra opção ao teste de qui-quadrado de Pearson é o teste de razão de verossimilhanças. Para a obtenção desse teste, deve-se substituir o estimador de máxima verossimilhança ($\hat{\pi}$) na expressão (2.3), para se obter o máximo da função de verossimilhança (Ferreira, 2008).

Assim, o máximo da função de verossimilhança irrestrita é dado por:

$$L_{\Omega}(\mathbf{y}; \hat{\pi}) = \frac{n!}{n_1!n_2! \dots n_k!} \prod_{i=1}^k \left(\frac{n_i}{n}\right)^{n_i}.$$

Sob a hipótese nula, nenhum parâmetro é estimado. Portanto, o máximo da função de verossimilhança é:

$$L_{\Omega_0}(\mathbf{y}; \pi_0) = \frac{n!}{n_1!n_2! \dots n_k!} \prod_{i=1}^k \pi_{0i}^{n_i}.$$

Assim, a estatística de razão de verossimilhanças é dada por:

$$\Lambda = \frac{L_{\Omega_0}(\mathbf{y}; \pi_0)}{L_{\Omega}(\mathbf{y}; \hat{\pi})} = \prod_{i=1}^k \frac{\pi_{0i}^{n_i}}{\left(\frac{n_i}{n}\right)^{n_i}},$$

que permite obter $-2\ln(\Lambda)$ por:

$$\chi_2^2 = 2 \sum_{i=1}^k n_i \ln \left(\frac{n_i}{n\pi_{0i}} \right)$$

que possui distribuição assintoticamente qui-quadrado com $\nu = k-1$ graus de liberdade sob H_0 .

2.8 Testes de distribuições baseadas em formas quadráticas

O conhecimento das formas quadráticas é de grande importância na estatística multivariada, pois muitos métodos de estimação e inferência são baseados em distâncias e na suposição de que os dados têm distribuição normal multivariada. As distâncias quadráticas e a densidade normal multivariada podem ser expressas em termos de matriz produto, denominadas de formas quadráticas (Johnson & Wichern, 2002).

Por definição, se A é uma matriz simétrica, $k \times k$ e \mathbf{X} e \mathbf{Y} são dois vetores em \mathbb{R}^p de dimensão $1 \times k$. Então, as expressões para a distância quadrática entre os pontos \mathbf{X} e \mathbf{Y} e para as formas quadráticas são dadas, respectivamente, por: $d^2(\mathbf{x}, \mathbf{y}) = (\mathbf{x} - \mathbf{y})^\top A (\mathbf{x} - \mathbf{y})$ e $Q(\mathbf{x}) = \mathbf{x}^\top A \mathbf{x} = \sum_{i=1}^k a_{ii} x_i^2 + \sum_{i=1}^{k-1} \sum_{j=i+1}^k a_{ij} x_i x_j = \sum_{i=1}^k \sum_{j=1}^k a_{ij} x_i x_j$.

De acordo com Ferreira (2008), é possível realizar inferências dos principais parâmetros de interesse utilizando a distribuição de formas quadráticas derivadas de vetores aleatórios de distribuições amostrais multivariadas. Existem alguns teoremas que relacionam a distribuição de formas quadráticas com os vetores de médias e matrizes de covariâncias de distribuições normais.

Um desses importantes teoremas, enunciado a seguir, afirma que:

Teorema 1: Se o vetor aleatório $\mathbf{Y} \in \mathbb{R}^p$ segue uma distribuição normal multivariada com densidade $f_{\mathbf{Y}}(\mathbf{y}) = (2\pi)^{-\frac{k}{2}} |\Sigma|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\mathbf{y} - \boldsymbol{\mu})^\top \Sigma^{-1} (\mathbf{y} - \boldsymbol{\mu}) \right\}$, com média

$\boldsymbol{\mu}$ e matriz de covariâncias $\boldsymbol{\Sigma}$, então:

$$(\mathbf{Y} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{Y} - \boldsymbol{\mu})$$

tem distribuição qui-quadrado com k graus de liberdade (χ_k^2) e a região

$$(\mathbf{Y} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{Y} - \boldsymbol{\mu}) \leq \chi_{\alpha, k}^2$$

define o elipsóide de concentração $(1 - \alpha) \times 100\%$ das realizações das variáveis aleatórias, em que $\chi_{\alpha, k}^2$ é o quantil superior $100\alpha\%$ da distribuição de qui-quadrado com $\nu = k$ graus de liberdade que, probabilisticamente, implica em $P(\chi_k^2 > \chi_{\alpha, k}^2) = \alpha$.

A região de confiança (RC) representa um subconjunto aleatório de \mathbb{R}^p que garante uma confiança de $100(1 - \alpha)\%$ de que esta região contenha $\boldsymbol{\mu}$. Para o caso particular de matriz de covariâncias $\boldsymbol{\Sigma}$ conhecida, a região de confiança para a média de uma população normal multivariada pode ser construída utilizando-se o teorema 1. Assim, sabendo que o ponto de massa $\bar{\mathbf{Y}}$ possui distribuição normal multivariada com vetor de médias $\boldsymbol{\mu}$ e matriz de covariâncias $\boldsymbol{\Sigma}/n$, quando a amostra aleatória é obtida de uma população normal multivariada, a região de $100(1 - \alpha)\%$ de confiança para o vetor de médias $\boldsymbol{\mu}$ é dada por:

$$\text{RC} = \left\{ \boldsymbol{\mu} \in \mathbb{R}^p \mid n(\boldsymbol{\mu} - \bar{\mathbf{Y}})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu} - \bar{\mathbf{Y}}) \leq \chi_{k, \alpha}^2 \right\}$$

Essa região é determinada pela hiperelipsóide de distância quadrática constante $\chi_{k, \alpha}^2/n$ do centro de massa definido por $\bar{\mathbf{Y}}$.

A construção da região de confiança de forma gráfica pode ser obtida apenas quando o número de variáveis é no máximo igual a três. Nesse caso, é possível identificar visualmente as variáveis responsáveis pela rejeição da hipótese nula $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$. Para dimensões maiores, pode-se observar se um determinado ponto pertence

ou não à região de confiança de forma analítica, se a desigualdade $n(\boldsymbol{\mu} - \bar{\mathbf{Y}}.)^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu} - \bar{\mathbf{Y}}.) \leq \chi_{k, \alpha}^2$ se verifica (Ferreira, 2008).

Outro procedimento alternativo que pode ser utilizado para identificar as variáveis que levaram à rejeição da hipótese nula são os intervalos de confiança para cada componente ou para combinações lineares dos componentes do vetor de médias. No entanto, para garantir que os intervalos preservem simultaneamente o coeficiente de confiança no nível nominal determinado, intervalos de confiança simultâneos são construídos.

Desse modo, os intervalos de $100(1 - \alpha)\%$ de confiança simultâneos com variâncias conhecidas são estabelecidos por:

$$\text{IC}_{(1-\alpha)}(\boldsymbol{\ell}^\top \boldsymbol{\mu}) : \boldsymbol{\ell}^\top \bar{\mathbf{Y}}. \pm \sqrt{\chi_{\alpha, k}^2} \sqrt{\frac{\boldsymbol{\ell}^\top \boldsymbol{\Sigma} \boldsymbol{\ell}}{n}}$$

em que $\boldsymbol{\ell}$ é um vetor não nulo de coeficientes reais utilizados para estabelecer a combinação linear de interesse.

Johnson & Weerahandi (1988) obtiveram regiões de credibilidade para a diferença de duas médias normais multivariadas utilizando formas quadráticas, mostrando uma solução ao problema de Behrens-Fisher multivariado. Um exemplo numérico foi apresentado para ilustrar o procedimento proposto.

Em estudos envolvendo proporções binomiais, o interesse está em realizar inferências sobre os parâmetros desconhecidos $\boldsymbol{\pi}$. Nesses casos, a normalidade não existe e pode ser obtida apenas de maneira aproximada, utilizando-se o teorema do Limite Central, que garante, para grandes valores de n , que $\hat{\boldsymbol{\pi}}$ possui distribuição aproximadamente normal multivariada, ou seja, $\hat{\boldsymbol{\pi}} \sim N(\boldsymbol{\pi}, \boldsymbol{\Sigma})$. Portanto, é possível inferir sobre proporções binomiais utilizando a aproximação normal assintótica e o Teorema 1 com as modificações necessárias.

2.9 Aplicações testes assintóticos

Motivado por problemas práticos de análise de dados *in vitro* envolvendo ensaios de anomalias cromossômicas, Williams (1988) estabeleceu vários testes estatísticos para verificar diferenças entre várias proporções binomiais. Dentre esses testes, foram avaliados o teste de qui-quadrado de Pearson, na presença de uma pequena correção (X^2) e o teste de razão de verossimilhanças (G^2). O autor verificou que o teste X^2 apresentou desempenho melhor do que o teste G^2 , uma vez que o tamanho desse teste se aproximou mais do nível nominal adotado de 5%. Os testes X^2 e G^2 foram conservativos, principalmente quando o número de sucessos era pequeno. Nas situações em que o número de sucessos eram maiores, o teste G^2 foi liberal. No entanto, verificou-se que essa liberalidade reduziu substancialmente ao aplicar um fator de ajustamento de Bartlett nesse teste.

Hayter & Liu (1990) consideraram o problema de testar a igualdade de várias probabilidades de Bernoulli, com a atenção voltada para a avaliação das propriedades de poder de um teste baseado na distância da transformação da raiz arco seno das proporções observadas. Os autores mostraram como é possível calcular os valores de poder sob a suposição assintótica e exata e comentaram que o teste proposto apresenta como vantagens, considerando grandes amostras, a obtenção de pelo menos uma configuração favorável de probabilidades e também a possibilidade de se obter uma expressão conveniente para avaliar a função poder exata.

Kulkarni & Shah (1995) desenvolveram um teste de igualdade de várias proporções binomiais em relação a um padrão conhecido. Para testar tal hipótese contra uma hipótese alternativa bilateral, esses autores estudaram a distribuição não nula da estatística teste sob a hipótese alternativa e estabeleceram uma maneira de calcular o poder via o método aproximado de momentos. Eles propuseram também uma estatística teste para testar a hipótese de igualdade das proporções contra uma hipótese alternativa uni-

lateral e apresentaram o método de obtenção dos valores-p e do poder desta estatística. Um método exato, utilizando somas ponderadas de distribuições de qui-quadrado com binomiais, foi avaliado numericamente em relação ao método aproximado proposto, por meio de dois exemplos de dados balanceados e não-balanceados.

McCulloch & Searle (2001) apresentaram alguns procedimentos que podem ser utilizados para inferir a respeito de duas ou mais proporções binomiais. Para testar a hipótese de igualdade de várias proporções binomiais, eles utilizaram os testes de razão de verossimilhanças e o de independência de qui-quadrado, que assintoticamente baseiam-se em uma distribuição de qui-quadrado. Considerando ainda grandes amostras, os autores apresentaram testes de hipóteses e intervalos de confiança que podem ser utilizados para comparar duas proporções binomiais e que são fundamentados em aproximações normais padrão. No entanto, segundo Hochberg & Tamhane (1987), esses métodos não preservam o valor global da significância, quando as comparações das diferenças das proporções, tomadas duas a duas, são realizadas.

Em virtude desse fato, McCulloch & Searle (2001) propuseram procedimentos alternativos que podem ser utilizados para analisar dados binomiais, dentre os quais destacam-se os modelos lineares generalizados. A teoria dos modelos lineares generalizados propicia uma melhor tratabilidade dos dados, pois permite que a distribuição da variável resposta seja qualquer distribuição da família exponencial, que inclui o modelo binomial. A estimação dos parâmetros desses modelos, geralmente, é realizada pelo método da máxima verossimilhança e as inferências baseiam-se na teoria assintótica.

Agresti (2002) mostrou que os testes de qui-quadrado e razão de verossimilhanças podem ser utilizados para testar a igualdade de proporções binomiais, no contexto de tabelas de contingência. Nessas condições, foram apresentados alguns exemplos para ilustrar a aplicação desses testes.

Krishnamoorthy et al. (2004) afirmam que estudos numéricos comprovaram que o tamanho do teste aproximado de Kulkarni & Shah (1995) frequentemente excedem

consideravelmente o nível nominal de significância. Por esta razão, esses autores apresentaram um método exato baseado na estatística teste de Kulkarni & Shah (1995), para testar a igualdade de várias proporções binomiais em relação a um padrão especificado e concluíram que, no caso de uma amostra, não existe nenhuma evidência vantajosa entre o teste proposto e o teste exato usual. Eles apresentaram também um procedimento de construção de intervalo de confiança para identificar as proporções populacionais que causaram a rejeição da hipótese de nulidade. Para ilustração dos métodos foram utilizados os exemplos apresentados por Kulkarni & Shah (1995).

Krishnamoorthy & Peng (2008) avaliaram o desempenho de alguns testes de igualdade de várias proporções binomiais. A distribuição binomial exata foi utilizada para avaliar as taxas de erro tipo I dos seguintes testes: condicional exato, condicional baseado no valor-p médio, qui-quadrado usual e o incondicional aproximado, denominado de UA-teste, proposto nesse trabalho. O UA-teste e o teste condicional controlaram as taxas de erro tipo I satisfatoriamente para pequenas amostras, enquanto o teste condicional exato foi considerado conservativo. Comparando-se os valores de poder desses três últimos testes, nas situações em que eles controlaram as taxas de erro tipo I, os autores verificaram que todos os testes apresentaram propriedades de poder similares. Conclui-se também que o AU-teste comporta-se praticamente como um teste exato, mesmo para pequenas amostras e pode ser usado com segurança para aplicações. Esses resultados foram ilustrados utilizando exemplos em que as proporções binomiais eram pequenas.

2.10 Simulação Monte Carlo

Em estudos de desempenho de testes de comparações múltiplas é bastante complicado obter analiticamente informações sobre as taxas de erro tipo I e poder dos testes. Também é impossível desenvolver pesquisas com dados reais, pois é difícil estabele-

cer vários experimentos sob as mesmas condições experimentais (Carmer & Swanson, 1973). Em virtude disso, tornou-se necessário utilizar os métodos de simulação, que permitem obter resultados sob situações desejáveis. Um método de simulação de grande aplicação nesse tipo de estudo é o Monte Carlo, que permite obter resultados de maneira mais simples e eficiente (Smith & Gelfand, 1992).

O método de Monte Carlo é um processo que consiste em simular dados por meio de um algoritmo, estabelecido em uma certa linguagem de programação, para gerar números pseudoaleatórios, com uma determinada distribuição de probabilidade. A finalidade desse método é estudar o comportamento de diferentes técnicas estatísticas que podem ser empregadas num problema específico (Dachs, 1988).

Na inferência clássica, vários autores aplicaram simulação Monte Carlo para comparar diversos testes de comparações múltiplas, levando em consideração as taxas de erro e poder. Entre eles, podem-se citar Carmer & Swanson (1973), Piegorsch (1990), Borges & Ferreira (2003) e Tsai et al. (2003), entre outros. Agresti & Min (2005), avaliando o desempenho de intervalo de credibilidade bayesiano em tabelas de contingência 2×2 para comparar proporções binomiais, também ressaltaram a importância de se usar os métodos de simulação e recomendaram utilizar o método de Monte Carlo para validar testes bayesianos.

3 Referências Bibliográficas

- AGRESTI, A. **Categorical data analysis**. New York: J. Wiley, 2002. 734p.
- AGRESTI, A.; BINI, M.; BERTACCINI, B.; RYU, E. Simultaneous confidence intervals for comparing binomial parameters. **Biometrics**, Washington, v.68, n.4, p.1-6, Dec. 2008.
- AGRESTI, A.; COULL, B.A. Approximate is better than “exact” for interval estimation of binomial proportions. **American Statistician**, Alexandria, v.52, n.2, p.119-126, May 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, June 2005.
- ALI, M.M.; CHO, J.S.; BEGUM, M. Bayesian multiple comparisons with nonparametric Dirichlet process priors for negative binomial populations. **Pak. Journal Statistical**, Lahore, v.22, n.2, p.89-98, 2006.
- ANDRADE, P.C.R. **Comparações múltiplas bayesianas em modelos normais homocedásticos e heterocedásticos**. 2008. 96p. Tese (Doutorado em Estatística) - Universidade Federal de Lavras, Lavras.
- 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, Dec. 1987.
- BENJAMINI, Y.; HOCHBERG, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. **Journal of the Royal Statistical Society**, London, v.57, n.1, p.289-300, 1995.
- BENJAMINI, Y.; LIU, W. A setp down multiple hypotheses testing procedure that controls the false discovery rate under independence. **Journal of Statistical Planning and Inference**, Amsterdam, v.82, n.1-2, p.163-170, 1999.
- BERRY, D.A.; HOCHBERG, Y. Bayesian perspectives on multiple comparisons. **Journal of Statistical Planning and Inference**, Amsterdam, v.82, n.1-2, p.215-227, 1999.
- BIASE, N.G. **Comparações múltiplas para parâmetros binomiais utilizando bootstrap**. 2006. 73p. Dissertação (Mestrado em Agronomia) - Universidade Federal de Lavras, Lavras.

- BIASE, N.G.; FERREIRA, D.F. Comparações múltiplas para proporções binomiais utilizando bootstrap. **Revista de Matemática e Estatística**, São Paulo, v.24, n.1, p.95-112, 2006.
- BOLFARINE, H.; SANDOVAL, M.C. **Introdução à inferência estatística**. Rio de Janeiro: SBM, 2000, 125p.
- BOLSTAD, W.M. **Introduction to bayesian statistics**. New York: J. Wiley, 2004. 354p.
- 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. **Revista de Matemática e Estatística**, São Paulo, v.21, n.1, p.67-83, 2003.
- BOX, G.E.P.; TIAO, G.C. **Bayesian inference in statistical analysis**. London: J. Wiley, 1992, 603p.
- BRATCHER, T.; HAMILTON, C. A bayesian multiple comparison procedure for ranking the means of normally distributed data. **Journal of Statistical Planning and Inference**, Amsterdam, v.133, n.1, p.23-32, 2005.
- BROEMILING, L.D. **Bayesian analysis of linear models**. New York: J. Wiley, 1989, 412p.
- CARMER, S.G.; SWANSON, M.R. An evaluation of ten pairwise multiple comparison procedures by Monte Carlo methods. **Journal American Statistical Association**, Washington, v.68, n.341, p.66-74, Mar. 1973.
- CHIB, S.; GREENBERG, E. Understanding the Metropolis-Hastings algorithm. **The American Statistician**, Salt Lake, v.49, n.4, p.327-345, Nov. 1995.
- CONLON, M.; THOMAS, R.G. A new confidence interval for the difference of two binomial proportions. **Computational Statistics & Data Analysis**, Amsterdam, v.9, n.2, p.237-241, Mar. 1990.
- CONSONNI, G.; VERONESE, P. A bayesian method for combining results from several binomial experiments. **Journal American Statistical Association**, New York, v.90, n.431, p.935-944, 1995.
- DACHS, J.N. **Estatística computacional: uma introdução em turbo Pascal**. Rio de Janeiro: Livros Técnicos e Científicos, 1988. 236p.
- DOBSON, A.J.; BARNETT, A. **An introduction to generalized linear models**. 3. ed. London: Chapman and Hall, 2008, 244p.

- DUNCAN, D.B. A bayesian approach to multiple comparisons. **Technometrics**, Washington, v.7, p.171-222, 1965.
- FERREIRA, D.F.; DEMÉTRIO, C.G.B.; MANLY, B.F.J.; MACHADO, A.A. Aplicações dos métodos de bootstrap nos procedimentos de comparações múltiplas. In: REUNIÃO ANUAL DA REGIÃO BRASILEIRA DA SOCIEDADE INTERNACIONAL DE BIOMETRIA, 50., 2005, Londrina. **Anais...** Londrina: SEAGRO, 2005, 5p.
- FERREIRA, D.F. **Estatística básica**. Lavras: UFLA, 2005. 664p.
- FERREIRA, D.F. **Estatística multivariada**. Lavras: UFLA, 2008. 662p. 1 v.
- GAMERMAN, D. **Markov Chain Monte Carlo**: stochastic simulation for bayesian inference. London: Chapman and Hall, 1997. 245p.
- GELFAND, A.E. Gibbs sampling. **Journal of the American Statistical Association**, London, v.95, p.1300-1304, 2000.
- GELFAND, A.E.; SMITH, A.F.M. Sampling based approaches for calculating maginal densities. **Journal of the American Statistical Association**, New York, v.85, n.410, p.398-409, June 1990.
- GELMAN, A.; CARLIN, J.B.; STERN, I.S.; RUBIN, D.B. **Bayesian data analysis**. London: Chapman and Hall, 1997. 668p.
- GELMAN, A.; RUBIN, D.B. Inference from iterative simulation using multiple sequences. **Statistical Science**, Hayward, v.7, n.4, p.457-511, 1992.
- GELMAN, A.; GELMAN, D. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. **IEEE Transactions on pattern Analysis and Machine intelligence**, Los Alamitos, v.6, n.6, p.721-741, 1984.
- GEWEKE, J. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In: BERNARDO, J.M.; BERGER, J.O.; DAWID, A.P.; SMITH, A.F.M. (Ed). **Bayesian statistics 4**. New Work: Oxford University, p.625-631, 1992.
- GOODMAN, L.A. Simultaneous confidence intervals for contrasts among multinomial populations. **Annals Mathematics Statistics**, v.35, n.2, p.716-725, 1964.
- GOPALAN, R.; BERRY, D.A. Bayesian multiple comparisons using Dirichlet process priors. **Journal of the American Statistical Association**, New York, v.93, n.443, p.1130-1139, 1998.

- HAYTER, A.J.; LIU, W. Power assessment for tests of the equality of several proportions. **Communications in Statistics. Theory and Methods**, New York, v.19, n.1, p.19-30, 1990.
- HEIDELBBER, G.P.; WELCH, P. Simulation run length control in the presence of an initial transient. **Operations Research**, Landing, v.31, n.6, p.1109-1144, Nov./Dec. 1983.
- HINKELMANN, K.; KEMPTHORNE, O. **Design and analysis of experiments**. New York: J. Wiley, 1987. 445p. 1 v.
- HOCHBERG, Y.; TAMHANE, A.C. **Multiple comparison procedures**. New York: J. Wiley, 1987. 450p.
- HSU, J.C. **Multiple comparisons: theory and methods**. London: Chapman and Hall, 1996, 277p.
- JEFFREYS, H. **Theory of probability**. London: Oxford University, 1961. 325p.
- JHUN, M.; JEONG, H.C. Applications of bootstrap methods for categorical data analysis. **Computational Statistics & Data Analysis**, Amsterdam, v.35, n.1, p.83-91, 2000.
- JOHNSON, R.A.; WEERAHANDI, S. A bayesian solution to the multivariate Behrens-Fisher problem. **Journal of the American Statistical Association**, New York, v.83, n.401, p.145-149, 1988.
- JOHNSON, R.A.; WICHERN, D.W. **Applied multivariate statistical analysis**. 3. ed. Englewood Cliffs: Prentice-Hall, 2002. 642p.
- KRISHNAMOORTHY, K.; PENG, J. Exact properties of a new test and other tests for differences between several binomial proportions. **Journal of Applied Statistical Science**, 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. **Computational Statistics & Data Analysis**, Amsterdam, v.45, p.697-707, 2004.
- KIM, B.S.; ZHAO, B.; KIM, H.J.; CHO, M.H. The statistical analysis of the in vitro chromosome aberration assay using Chinese hamster ovary cells. **Mutation research-genetic toxicology and environmental mutagenesis**, v.469, p.243-252, June 2000.

- KULKARNI, P.M.; SHAH, A. Testing the equality of several binomial proportions to a prespecified standard. **Statistics & Probability Letters**, Amsterdam, v.25, n.3, p.213-219, 1995.
- MACHADO, A.A.; DEMÉTRIO, C.G.B.; FERREIRA, D.F.; SILVA, J.G.C. Estatística experimental: uma abordagem fundamentada no planejamento e no uso de recursos computacionais. In: REUNIÃO ANUAL DA REGIÃO BRASILEIRA DA SOCIEDADE INTERNACIONAL DE BIOMETRIA, 50., 2005, Londrina. **Anais...** Londrina: ISBN, 2005. 290p.
- MARDIA, K.V.; KENT, J.T.; BIBBY, J.M. **Multivariate analysis**. London: Academic, 1995. 518p.
- MCCANN, M.H.; TEBBS, J.M. Pairwise comparisons for proportions estimated by pooled testing. **Journal of Statistical Planning and Inference**, Amsterdam, v.137, n.4, p.1278-1290, 2007.
- MCCULLOCH, C.E.; SEARLE, S.R. **Generalized, linear, and mixed models**. New York: J. Wiley, 2001. 346p.
- MOOD, A.M.; GRAYBILL, F.A.; BOES, D.C. **Introduction to the theory of statistics**. 3. ed. New York: J. Wiley, 1974. 564p.
- NELDER, J.A.; WEDDERBURN, R.W.M. Generalized linear models. **Journal of the Royal Statistical Society**, London, v.135, p.370-384, 1972.
- O'NEILL, R.; WETHERILL, G.B. The present state os multiple comparison methods. **Journal of the Royal Statistical Society**, Oxford, v.33, n.2. p.218-250, 1971.
- PAN, W. Approximate confidence intervals for one proportion and two proportions. **Computational Statistics & Data Analysis**, Amsterdam, v.40, n.1, p.143-157, July 2002.
- PAULINO, C.D.; TURKMAN, M.A.A.; MURTEIRA, B. **Estatística bayesiana**. Lisboa: Fundação Calouste Gulbenkian, 2003. 444p.
- PERECIN, D.; BARBOSA, J.C. Uma avaliação de seis procedimentos para comparações múltiplas. **Revista de Matemática e Estatística**, Jaboticabal, v.6, p.95-103, 1988.
- PERRY, J.N. Multiple comparison procedures: a dissenting view. **Journal of Economic Entomology**, Menasha, v.79, n.5, p.1149-1155, 1986.
- PIEGORSCH, W.W. Multiple comparisons for analysing dichotomous response. **Biometrics**, Washington, v.47, n.1, p.45-52, Mar. 1990.

RAFTER, J.A.; ABELL, M.L.; BRASELTON, J.P. Multiple comparison methods for means. **SIAM Review**, Philadelphia, v.44, n.2, p.259-278, 2002.

RAFTERY, A.L.; LEWIS, S. Comment: one long run with diagnostics: implementation strategies for Markov chain Monte Carlo. **Statistical Science**, Hayward, v.7, n.4, p.493-497, 1992.

RAMALHO, M.A.P.; FERREIRA, D.F.; OLIVEIRA, A.C. **Experimentação em genética e melhoramento de plantas**. Lavras: UFLA, 2000. 303p.

RAVEN, P.H.; EVERT, R.F.; EICHHOR, S.E. **Biology of plants**. 7. ed. New York: Freeman and Company, 2005. 960p.

R Development Core Team. **R: A language and environment for statistical computing**. Vienna: R Foundation for Statistical Computing, 2008. Disponível em: <<http://www.r-project.org>>. Acesso em: 10 out. 2008.

RINGLAND, J.T. Robust multiple comparisons. **Journal of the American Statistical Association**, New York, v.78, p.145-151, 1983.

SANTOS, C.; FERREIRA, D.F.; BUENO FILHO, J.S.S. Novas alternativas de testes de agrupamento avaliadas por meio de simulação Monte Carlo. **Ciência e Agrotecnologia**, Lavras, v.25, n.6, p.1382-1392, nov. 2001.

SCHEFFÉ, H. A method for judging all contrasts in the analysis of variance. **Biometrika**, London, v.40, n.1-2, p.87-104, 1953.

SCOTT, G.J.; BERGER O.J. An exploration of aspects of Bayesian multiple testing. **Journal of Statistical Planning and Inference**, Amsterdam, v.136, n.7, p.2144-2162, July 2006.

SILVA E.C.; FERREIRA, D.F.; BEARZOTI, E. Avaliação do poder e taxas de erro tipo I do teste de Scott-Knott por meio do método de Monte Carlo. **Ciência e Agrotecnologia**, Lavras, v.23, n.3, p.687-696, jul./set. 1999.

SMITH, C.W.; GELFAND, A.E. Bayesian statistics without tears: a sampling resampling perspective. **The American Statistician**, Alexandria, v.46, p.84-88, May 1992.

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

STOREY, J.D. A direct approach to false discovery rates. **Journal of the Royal Statistical Society. Series B**, v.64, n.3, p.479-498, 2002.

TSAI, C.; HSUEH, H.; CHEN, J.J. Estimation of false discovery rates in multiple testing: application to gene microarray data. **Biometrics**, Washington, v.59, n.4, p.1071-1081, Dec. 2003.

WALLER, R.A.; DUNCAN, D.B. A bayes rule for the symmetric multiple comparisons problem. **Journal of the American Statistical Association**, New York, v.64, n.328, p.1484-1503, 1969.

WALPOLE, R.E. **Introduction to statistics**. New York: J. Wiley, 1974. 340p.

WESTFALL, P.H.; YOUNG, S.S. P-value adjustments for multiple tests in multivariate binomial models. **Journal of the American Statistical Association**, New York, v.84, n.407, p.780-785, 1989.

WILLIAMS, D.A. Test for differences between several small proportions. **Journal of the Royal Statistical Society**, London, v.37, n.3, p.421-434, 1988.

CAPÍTULO 2

COMPARAÇÕES MÚLTIPLAS E TESTE SIMULTÂNEO PARA PARÂMETROS BINOMIAIS INDEPENDENTES DE k POPULAÇÕES

1 RESUMO

Uma estratégia utilizada em estatística para comparar várias proporções binomiais é a aplicação do teste F em uma análise de variância, seguida de testes de comparações múltiplas, se houver a rejeição da hipótese nula global. No entanto, as pressuposições exigidas pelos testes nem sempre são atendidas e, portanto, essa estratégia não é adequada. Entre os métodos alternativos utilizados para contornar o problema, destacam-se os testes assintóticos. Este trabalho teve por objetivos propor um teste de comparações múltiplas assintóticas de proporções binomiais, baseado em formas quadráticas e avaliar seu desempenho por meio de simulação Monte Carlo, além de divulgar e avaliar o desempenho dos testes assintóticos G^2 e X^2 de Pearson para a hipótese de várias proporções binomiais. Foram simuladas 10.000 amostras de Monte Carlo para diferentes configurações de k populações binomiais geradas com parâmetros π_i e n_i para a i -ésima população, $i = 1, 2, \dots, k$. 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. Sob H_0 completa, foram consideradas populações com π_i 's idênticos e fixados em 0,01, 0,1 e 0,5 e, sob H_0 parcial, considerou-se uma diferença Δ , variando de 0,01 a 0,9, entre os valores de π de dois grupos distintos. Foram consideradas $k = 2, 5$ e 10 populações e tamanhos de amostras $n_i = 10, 30$ e 100 para cada população, admitindo-se os níveis nominais de significância de 1% e 5%. De modo geral, o teste X^2 controlou o erro tipo I, em níveis iguais ou inferiores aos valores nominais de significância, e apresentou desempenho superior ao do teste G^2 , que foi considerado liberal, principalmente nas situações de pequenas amostras e maiores populações. O teste de comparações múltiplas assintóticas proposto apresentou excelentes resultados e, 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.

Palavras-chave: Simulação Monte Carlo, razão de verossimilhanças, formas quadráticas, procedimentos de comparações múltiplas.

2 ABSTRACT

A strategy used in statistics 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 also 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 10.000 Monte Carlo samples for different configuration of k binomial populations generated with parameters π_i and n_i for the i th population, $i = 1, 2, \dots, k$. For the likelihood ratio and chi-square based on quadratic forms tests the type I error rates were measured only under H_0 complete. For the specific case of multiple comparisons the experimentwise type I error rates were computed under complete and partial H_0 . The power of all the tests was assessed under partial H_0 . Under complete H_0 populations with identical π_i 's settled in 0.01, 0.1 and 0.5 were considered and under partial H_0 two different groups were considered with a difference Δ , varying from 0.01 to 0.9, between the values of π of the two groups. Several number of populations, $k = 2, 5$ and 10, and several sample sizes, $n_i = 10, 30$ and 100, were considered in the Monte Carlo simulations assuming the nominal levels of significance of 1% and 5%. In general, 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 considered liberal, especially for the small samples and large number of populations circumstances. The asymptotic multiple comparisons test proposed showed excellent results, and since the experimentwise type I error rates were conservative and the power was high this test is recommended for real situations.

Key-words: Monte Carlo simulation, likelihood ratio, quadratic forms, multiple comparison procedures.

3 Introdução

Em diferentes áreas do conhecimento, o pesquisador se depara com a necessidade de realizar inferências a respeito de várias proporções binomiais. Esse é o caso, por exemplo, dos ensaios de germinação (Raven et al., 2005) e dos ensaios de ebsenação cromossômica (Williams, 1988; Kim et al., 2000). 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, 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 & Searle, 2001). Nessa alternativa, a família exponencial, que abrange o modelo binomial, é utilizada. Assim, a distribuição adequada dos dados é empregada, embora a distribuição das estatísticas sejam apenas assintóticas (Nelder & Wedderburn, 1972; Dobson & 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 reposição (*bootstrap*) e sem reposição (permutação). Entre os estudos envolvendo proporções binomiais, destaca-se o trabalho de Biase & Ferreira (2006), no qual testes de comparações múltiplas foram propostos e avaliados os desempenhos desses testes. Os autores não apresentaram, no entanto, 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 & Peng (2008). Entre os testes destacam-se o da razão de verossimilhanças G^2 e o de X^2 de Pearson. Um outro teste para a igualdade de proporções binomiais a um valor de referência π_0 pré-determinado é apresentado por Krishnamoorthy et al. (2004).

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. Em uma segunda etapa, o autor considerou um método implementado por Hochberg & Tamhane (1987), usando o intervalo de Wald juntamente com a distribuição da amplitude normal padronizada, para comparar todas as combinações das proporções tomadas duas a duas. 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-se um procedimento de estimação intervalar simultâneo utilizando a formulação de Jeffreys-Perks, motivada pela aproximação bayesiana de Beal (1987).

Por essas razões, o presente trabalho foi realizado com os objetivos de 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.

4 Metodologia

Considerando amostras aleatórias independentes (eventos de Bernoulli) 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 π_0 não especificado, o que faz com que este trabalho difira do desenvolvido por Krishnamoorthy et al. (2004). O primeiro teste considerado refere-se ao teste da 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. A inclusão dos testes da hipótese $H_0: \pi_1 = \pi_2 = \dots = \pi_k = \pi_0$ neste capítulo teve o objetivo de obter resultados que servissem de referência para os resultados do teste bayesiano proposto no capítulo 3.

4.1 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. Para isso, a função de verossimilhança, considerando os modelos probabilísticos binomiais independentes, sob H_1 (modelo irrestrito Ω), é:

$$P(Y_1 = y_1, \dots, Y_k = y_k) = L_{\Omega}(\mathbf{y}|\boldsymbol{\pi}) = \prod_{i=1}^k \frac{n_i!}{y_i!(n_i - y_i)!} \pi_i^{y_i} (1 - \pi_i)^{n_i - y_i}, \quad (4.1)$$

em que $\mathbf{y} = [y_1, y_2, \dots, y_k]^T$ é o vetor de dados e $\boldsymbol{\pi} = [\pi_1, \pi_2, \dots, \pi_k]^T$.

Para o modelo restrito (Ω_0), sob H_0 , a função de verossimilhança é:

$$P(Y_1 = y_1, \dots, Y_k = y_k) = L_{\Omega_0}(\mathbf{y}|\pi_0) = \prod_{i=1}^k \frac{n_i!}{y_i!(n_i - y_i)!} \pi_0^{y_i} (1 - \pi_0)^{n_i - y_i}. \quad (4.2)$$

As funções de verossimilhanças (4.1) e (4.2) foram maximizadas e a estatística do teste resultante, na forma geral, é dada por:

$$\Lambda = \frac{L_{\Omega_0}(\mathbf{y}|\pi_0)}{L_{\Omega}(\mathbf{y}|\boldsymbol{\pi})}, \quad (4.3)$$

em que $\hat{\pi}_0$ é o estimador de máxima verossimilhança de π_0 no modelo restrito Ω_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 Ω , sob H_1 .

Foi utilizado o fato de que $\chi_{c1}^2 = -2\ln(\Lambda)$ 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 (4.1), foi obtida a função suporte, sendo considerado \mathbf{Y} como variável aleatória, resultando em:

$$\begin{aligned} g(\mathbf{Y}, \boldsymbol{\pi}) &= \sum_{i=1}^k \ln(n_i!) - \sum_{i=1}^k \ln(Y_i!) - \sum_{i=1}^k \ln[(n_i - Y_i)!] \\ &+ \sum_{i=1}^k Y_i \ln(\pi_i) + \sum_{i=1}^k (n_i - Y_i) \ln(1 - \pi_i). \end{aligned}$$

Derivando-se essa expressão em relação a π_i , tem-se:

$$\frac{\partial g(\mathbf{Y}, \boldsymbol{\pi})}{\partial \pi_i} = \frac{Y_i}{\pi_i} - \frac{n_i - Y_i}{1 - \pi_i},$$

que, igualando a zero, resulta no seguinte estimador de máxima verossimilhança:

$$\hat{\pi}_i = \frac{Y_i}{n_i}, \quad i = 1, 2, \dots, k. \quad (4.4)$$

Assim, o máximo da função de verossimilhança no modelo irrestrito é:

$$L_{\Omega}(\mathbf{Y}_i|\hat{\pi}_i) = \prod_{i=1}^k \frac{n_i!}{Y_i!(n_i - Y_i)!} \hat{\pi}_i^{Y_i} (1 - \hat{\pi}_i)^{n_i - Y_i}.$$

Para o modelo restrito, tem-se a seguinte função suporte:

$$\begin{aligned} g(\mathbf{Y}, \pi_0) &= \sum_{i=1}^k \ln(n_i!) - \sum_{i=1}^k \ln(Y_i!) - \sum_{i=1}^k \ln[(n_i - Y_i)!] \\ &+ \sum_{i=1}^k [Y_i \ln(\pi_0) + (n_i - Y_i) \ln(1 - \pi_0)], \end{aligned}$$

cuja derivada de primeira ordem em relação a π_0 é:

$$\frac{dg(Y, \pi_0)}{d\pi_0} = \frac{\sum_{i=1}^k Y_i}{\pi_0} - \frac{\sum_{i=1}^k (n_i - Y_i)}{1 - \pi_0}.$$

Igualando-se essa derivada a zero e resolvendo, tem-se o estimador de máxima verossimilhança para o modelo restrito (Ω_0) de π_0 dado por:

$$\hat{\pi}_0 = \frac{\sum_{i=1}^k Y_i}{n} = \frac{\sum_{i=1}^k n_i \hat{\pi}_i}{n}, \quad (4.5)$$

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 (4.4) de cada população, tendo como peso os tamanhos das amostras.

O máximo da função de verossimilhança restrita (4.2) é:

$$L_{\Omega_0}(\mathbf{Y}|\hat{\pi}_0) = \prod_{i=1}^k \frac{n_i!}{Y_i!(n_i - Y_i)!} \hat{\pi}_0^{Y_i} (1 - \hat{\pi}_0)^{n_i - Y_i}.$$

A estatística do teste, utilizando (4.3) é, portanto, dada por:

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

Assim, tem-se que:

$$\begin{aligned} \chi_{c1}^2 &= -2 \sum_{i=1}^k Y_i \ln \left(\frac{\hat{\pi}_0}{\hat{\pi}_i} \right) - 2 \sum_{i=1}^k (n_i - Y_i) \ln \left[\frac{1 - \hat{\pi}_0}{1 - \hat{\pi}_i} \right] \\ &= 2 \sum_{i=1}^k Y_i \ln \left(\frac{\hat{\pi}_0}{\hat{\pi}_i} \right)^{-1} + 2 \sum_{i=1}^k (n_i - Y_i) \ln \left[\frac{1 - \hat{\pi}_0}{1 - \hat{\pi}_i} \right]^{-1}. \end{aligned} \quad (4.6)$$

Substituindo-se os estimadores de π_i e π_0 , obtidos em (4.4) e (4.5), na expressão (4.6), obtém-se que:

$$\chi_{c1}^2 = 2 \left\{ \sum_{i=1}^k Y_i \ln \left(\frac{n Y_i}{n_i \sum_{j=1}^k Y_j} \right) + \sum_{i=1}^k (n_i - Y_i) \ln \left[\frac{n(n_i - Y_i)}{n_i(n - \sum_{j=1}^k Y_j)} \right] \right\} \quad (4.7)$$

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 $\chi_{c1}^2 \geq \chi_{\alpha, k-1}^2$, a hipótese de nulidade das proporções binomiais das k populações deve ser rejeitada. Esse teste 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.

4.2 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 π é assintoticamente normal com média π e matriz de covariância $\Sigma_{k \times k}$. Assim, utilizando-se os resultados de Ferreira (2008), verifica-se que:

$$\chi_{c2}^2 = (\hat{\pi} - \pi)^\top \Sigma^{-1} (\hat{\pi} - \pi) \quad (4.8)$$

possui distribuição qui-quadrado com k graus de liberdade.

Sob H_0 , o vetor π 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 Σ como alternativa. Portanto, a quantidade pivotal (4.8) utilizando tal estimador terá distribuição qui-quadrado assintótica com $\nu = k - 1$ graus de liberdade. A matriz Σ foi determinada a partir do valor hipotético π_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 $Y_i = \sum_{j=1}^{n_i} X_{ij}$. Essa variável aleatória X_{ij} assumirá o valor 1 para o sucesso do evento considerado e 0 para o fracasso. Para o desenvolvimento da expressão (4.8), foi necessário obter a matriz de covariâncias Σ que depende das expressões para as variâncias e covariâncias de Y_i . Assim,

$$\begin{aligned} E(X_{ij}) &= 1 \times P(X_{ij} = 1) + 0 \times P(X_{ij} = 0) \\ &= P(X_{ij} = 1) = \pi_i. \end{aligned}$$

Como a amostra de tamanho n_i da i -ésima população binomial é aleatória, então, $X_{i1}, X_{i2}, \dots, X_{in_i}$ são independentes. Logo,

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

Também se pode obter facilmente a esperança de X_{ij}^2 , pois X_{ij} assume valores iguais a 1 ou a 0. Portanto,

$$\begin{aligned} E(X_{ij}^2) &= 1^2 \times P(X_{ij} = 1) + 0^2 \times P(X_{ij} = 0) \\ &= P(X_{ij} = 1) = \pi_i. \end{aligned}$$

Assim,

$$\begin{aligned} V(X_{ij}) &= E(X_{ij}^2) - E^2(X_{ij}) \\ &= \pi_i^2 - \pi_i = \pi_i(1 - \pi_i). \end{aligned}$$

Portanto, a variância de Y_i é:

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

A covariância entre Y_i e $Y_{i'}$ para $i \neq i'$ é nula, uma vez que as amostras das diferentes populações são independentes. O objetivo é determinar a covariância de $\hat{\pi}$. Utilizando o estimador de máxima verossimilhança $\hat{\pi}_i$ (4.4), 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 Σ 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 à independência.

Dessa forma, a quantidade pivotal (4.8) sob H_0 pode ser simplificada em:

$$\chi_{c2}^2 = [\hat{\pi}_1 - \pi_0, \hat{\pi}_2 - \pi_0, \dots, \hat{\pi}_k - \pi_0] \begin{bmatrix} \frac{n_1}{\pi_1(1-\pi_1)} & \dots & 0 \\ 0 & \ddots & 0 \\ \vdots & \vdots & \vdots \\ 0 & \dots & \frac{n_k}{\pi_k(1-\pi_k)} \end{bmatrix} \begin{bmatrix} \hat{\pi}_1 - \pi_0 \\ \hat{\pi}_2 - \pi_0 \\ \vdots \\ \hat{\pi}_k - \pi_0 \end{bmatrix},$$

que, expandida, resulta em:

$$\chi_{c2}^2 = \sum_{i=1}^k \frac{n_i(\hat{\pi}_i - \pi_0)^2}{\pi_i(1-\pi_i)} = \sum_{i=1}^k \frac{(Y_i - n_i\pi_0)^2}{n_i\pi_i(1-\pi_i)}.$$

Essa quantidade pivotal não pode ser utilizada em situações reais, pois depende de π_i , que é desconhecido. Desse modo, uma alternativa que pode ser considerada baseia-se em um procedimento análogo ao χ^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 (4.5). Assim, a estatística desse teste é dada por:

$$\chi_{c2}^2 = \sum_{i=1}^k \frac{(Y_i - n_i\hat{\pi}_0)^2}{n_i\hat{\pi}_0(1-\hat{\pi}_0)}. \quad (4.9)$$

A estatística (4.9), sob H_0 , segue uma distribuição qui-quadrado assintótica com

$\nu = k - 1$ graus de liberdade. Os desempenhos das estatísticas (4.7) e (4.9) serão avaliados na seção (5.1).

4.3 Comparações múltiplas

Para realizar o teste de comparações múltiplas foi utilizado o resultado geral (4.8), devidamente modificado para contemplar o estimador $\hat{\pi}_0$ no lugar do parâmetro π_0 . Assim, para testar

$$H_0 : \pi_i = \pi_{i'}, \quad i \neq i' = 1, 2, \dots, k \quad (4.10)$$

foi utilizada a seguinte estatística:

$$\chi_{c3}^2 = \frac{(\hat{\pi}_i - \hat{\pi}_{i'})^2}{\text{var}(\hat{\pi}_i) + \text{var}(\hat{\pi}_{i'}) - 2\text{cov}(\hat{\pi}_i, \hat{\pi}_{i'})}, \quad (4.11)$$

que possui distribuição assintótica qui-quadrado com $\nu = k$ graus de liberdade, sendo $\text{var}(\hat{\pi}_i)$ e $\text{var}(\hat{\pi}_{i'})$ 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 (4.10) será rejeitada se $\chi_{c3}^2 > \chi_{\alpha, k}^2$, em que $\chi_{\alpha, k}^2$ é o quantil superior 100 α % da distribuição qui-quadrado com $\nu = k$ graus de liberdade. As variâncias e a covariância de (4.11) são desconhecidas e precisam ser estimadas. As particularidades desses estimadores estão apresentadas na seção 5.2.

4.4 Simulações Monte Carlo

Para a validação dos desempenhos dos testes avaliados neste trabalho foram realizadas simulações Monte Carlo. Foram geradas k populações binomiais com parâmetros π_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 amos-

tras 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 π_i 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 & 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 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 $\ell = 1, 2, \dots, 5$, 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 $\Delta = 0,01, 0,10$ 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.

No caso particular de $k = 5$ populações, a formação dos dois grupos foi realizada conforme Biase & Ferreira (2006), em que as populações 1, 2 e 3 pertenceram ao primeiro grupo e 4 e 5, ao segundo. As taxas de erro tipo I foram confrontadas com os valores obtidos por Biase & Ferreira (2006) utilizando *bootstrap* e comparadas com os níveis nominais de significância.

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 π 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 $N = 10000$ simulações Monte Carlo para o nível nominal α , então, a estatística 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 ν_1 e ν_2 graus de liberdade.

Para $N = 10.000$ simulações, verificou-se, para $\alpha = 5\%$, que valores inferiores a 445 e superiores a 557 levam à rejeição da hipótese nula $H_0 : \alpha = 5\%$, considerando o nível de significância de 1% para o teste. Da mesma forma, valores inferiores a 75 e superiores a 127 levam à rejeição da hipótese $H_0 : \alpha = 1\%$, considerando o mesmo nível de significância de 1% para o teste realizado.

Os valores de poder, no caso de comparações múltiplas, também foram comparadas com os valores apresentados em Biase & Ferreira (2006).

Para avaliação do desempenho dos testes propostos, foram implementadas as rotinas utilizando-se o programa R (R. Development Core Team, 2008), que são apresentadas em anexo.

5 Resultados e Discussão

5.1 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.

5.1.1 Erro tipo I sob H_0 completa

Na Tabela 1.2 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 π , resultantes de 10.000 simulações Monte Carlo, considerando o nível nominal de 5%.

Pode-se verificar, na maioria das 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 desempenho 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 desempenho 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.

Para o teste G^2 , houve um desempenho 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. Para $\pi = 0,1$, o teste foi liberal com $k = 2, 5$ e 10, considerando $n = 30$ e também com $k = 10$ e $n = 10$. Para $\pi = 0,5$, este teste foi liberal com $k = 2$ e $n = 100$

TABELA 1.2 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 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 ⁺	0,00 ⁺	4,66 ^{ns}	0,88 ⁺	5,21 ^{ns}	3,99 ⁺
2	30	0,45 ⁺	0,03 ⁺	7,22 [*]	5,40 ^{ns}	5,35 ^{ns}	5,33 ^{ns}
2	100	6,01 [*]	1,34 ⁺	5,02 ^{ns}	4,99 ^{ns}	5,68 [*]	5,68 [*]
5	10	0,04 ⁺	0,04 ⁺	5,12 ^{ns}	2,55 ⁺	6,49 [*]	4,33 ⁺
5	30	0,58 ⁺	0,53 ⁺	8,57 [*]	4,94 ^{ns}	5,63 [*]	5,12 ^{ns}
5	100	3,25 ⁺	2,66 ⁺	5,47 ^{ns}	4,85 ^{ns}	5,36 ^{ns}	5,31 ^{ns}
10	10	0,01 ⁺	1,78 ⁺	5,65 [*]	3,81 ⁺	7,29 [*]	4,75 ^{ns}
10	30	0,24 ⁺	4,36 ⁺	8,27 [*]	4,81 ^{ns}	6,17 [*]	5,55 ^{ns}
10	100	4,41 ⁺	4,35 ⁺	5,38 ^{ns}	4,51 ^{ns}	4,88 ^{ns}	4,71 ^{ns}

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

⁺ significativamente inferior ao nível nominal de 5%, considerando uma confiança de 99%.

^{ns} não significativamente diferente do nível nominal de 5%, considerando uma confiança de 99%.

e também com $k = 5$ e 10 , considerando $n = 10$ e 30 . Uma possível e potencial explicação para esse resultado pode ser dada em função da variância da binomial. Essa variância é máxima quando o parâmetro π é igual a $0,5$. Na medida que π afasta-se de $0,5$, o valor da variância diminui. Assim, espera-se que essa maior variabilidade influencie no desempenho do teste quanto a sua taxa de erro tipo I. 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 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 de esse 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).

Para o nível nominal de significância de 1%, os testes apresentaram taxas de erro tipo I mais conservativas do que as observadas para 5%. Esses resultados são apresentados na Tabela 1.3. Novamente ocorreu, especificamente, um caso em que o teste X^2 foi considerado liberal. Isso foi verificado para $k = 2$ e $n = 30$ com $\pi = 0,5$. Nas demais situações, as taxas de erro tipo I desse teste foram significativamente ($P < 0,01$) inferiores ou não diferiram ($P > 0,01$) significativamente do valor nominal de 1%. O teste X^2 foi conservativo em todas as situações avaliadas para $\pi = 0,01$, exceto com $k = 10$ e $n = 100$, em que o teste apresentou tamanho não significativamente ($P < 0,01$) diferente do valor nominal e também para $k = 2$ e $n = 10$ e 30 com $\pi = 0,1$. Para números de populações maiores ($k = 5$ e 10), o teste X^2 apresentou taxas de erro tipo I inferiores ao nível nominal de 1% somente para pequenas amostras ($n = 10$) e $\pi = 0,1$.

Pode-se observar que houve um melhor desempenho do teste X^2 para $\pi = 0,5$, pois todos os valores das taxas apresentaram tamanhos não significativamente ($P < 0,01$) diferentes do valor nominal de significância de 1%, com exceção para $k = 2$ e $n = 30$, em que o teste foi considerado liberal.

Considerando valores afastados de 0,5 ($\pi \leq 0,1$), o teste G^2 apresentou situações em que foi considerado conservativo, exato e liberal para o nível nominal de 1%. Todas as taxas de erro tipo I do teste G^2 , considerando $\pi = 0,01$, são significativamente ($P < 0,01$) inferiores ao valor nominal de 1%. Para $\pi = 0,1$, o teste G^2 apresentou desempenho conservativo para pequenas amostras ($n = 10$), independente do número de populações. Para esse mesmo valor de π , o teste foi considerado liberal para tamanhos de amostras intermediárias ($n = 30$) e exato para amostras grandes ($n = 100$).

TABELA 1.3 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 os testes de razão de verossimilhanças (G^2) e formas quadráticas (X^2), ao nível nominal de 1%.

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	0,00 ⁺	0,00 ⁺	0,13 ⁺	0,13 ⁺	1,15 ^{ns}	1,15 ^{ns}
2	30	0,00 ⁺	0,00 ⁺	1,72 [*]	0,66 ⁺	1,43 [*]	1,43 [*]
2	100	0,24 ⁺	0,00 ⁺	0,97 ^{ns}	0,84 ^{ns}	0,76 ^{ns}	0,76 ^{ns}
5	10	0,00 ⁺	0,00 ⁺	0,37 ⁺	0,49 ⁺	1,32 [*]	0,81 ^{ns}
5	30	0,01 ⁺	0,01 ⁺	1,74 [*]	0,89 ^{ns}	1,14 ^{ns}	0,99 ^{ns}
5	100	0,33 ⁺	0,38 ⁺	1,20 ^{ns}	0,91 ^{ns}	1,17 ^{ns}	1,14 ^{ns}
10	10	0,00 ⁺	0,13 ⁺	0,56 ⁺	0,62 ⁺	2,09 [*]	0,79 ^{ns}
10	30	0,01 ⁺	0,41 ⁺	1,79 [*]	1,01 ^{ns}	1,34 [*]	0,98 ^{ns}
10	100	0,46 ⁺	0,86 ^{ns}	1,06 ^{ns}	0,83 ^{ns}	1,10 ^{ns}	1,03 ^{ns}

* significativamente superior ao nível nominal de 1%, considerando uma confiança de 99%.

⁺ significativamente inferior ao nível nominal de 1%, considerando uma confiança de 99%.

^{ns} não significativamente diferente do nível nominal de 1%, considerando uma confiança de 99%.

O teste G^2 foi liberal para $\pi = 0,5$ com $k = 2$ e $n = 30$, $k = 5$ e $n = 10$ e também para $k = 10$ e $n = 10$ e 30. Em todas as demais situações, o tamanho do teste G^2 foi não significativamente diferente do valor nominal de 1%.

De modo geral, verificou-se que o teste G^2 apresentou melhores desempenhos, considerando grandes amostras, uma vez que, nessas situações, o teste foi exato.

Procurando avaliar o desempenho dos testes para tamanhos amostrais variando entre 10 e 100, foram feitas simulações adicionais, considerando $k = 5$. Nessas situações, o erro tipo I foi computado para os níveis nominais de significância de 1% e 5%.

Na Figura 1.1 são 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 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.

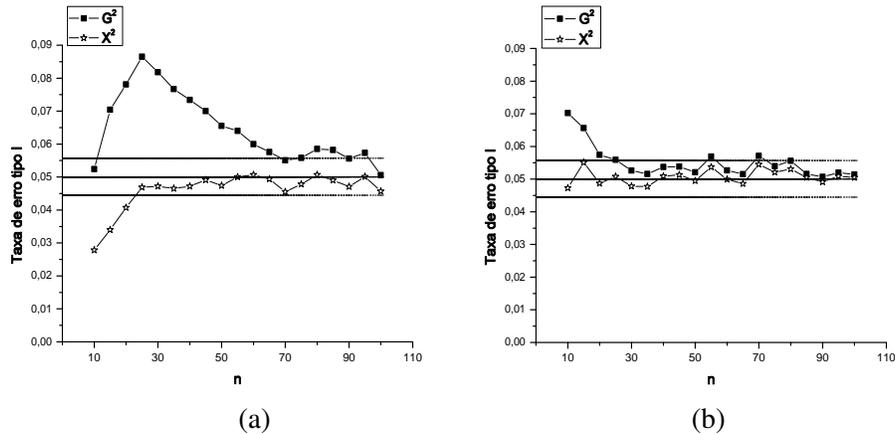


FIGURA 1.1 Taxas de erro tipo I dos testes de razão de verossimilhanças (G^2) e formas quadráticas (X^2), em função dos tamanhos amostrais (n) e valores dos parâmetros (a) $\pi = 0,1$ e (b) $\pi = 0,5$ para $\alpha = 5\%$, considerando a hipótese H_0 completa.

Para pequenas amostras ($n = 10$), na Figura 1.1 (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, as 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, pode-se observar, na Figura 1.1 (a), que o teste X^2 foi conservativo para amostras pequenas ($n \leq 20$) 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 1.1 (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 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 $\pi = 0,1$.

Para o valor nominal de significância de 1%, pode-se observar, na Figura 1.2 (b), que, para valores de $\pi = 0,5$, as taxas de erro tipo I dos dois testes apresentaram o mesmo desempenho geral dos testes obtidos para $\alpha = 5\%$. Para valores de π afastados de 0,5 ($\pi = 0,1$), os resultados das taxas do teste X^2 também foram bastante similares aos observados para 5%. No entanto, o valor da taxa de erro tipo I do teste G^2 foi significativamente menor do que 1% para pequenas amostras ($n = 10$), ou seja, o teste foi conservativo. Para $n = 15$ e 20, o tamanho do teste não diferiu significativamente ($P < 0,01$) do valor nominal e, para tamanhos de amostras variando de 25 a 60, os valores das taxas foram sempre superiores a 1%, indicando um desempenho liberal do teste. Para $n \geq 65$, as taxas do teste em questão foram consideradas iguais ou, em algumas situações superiores ao valor nominal.

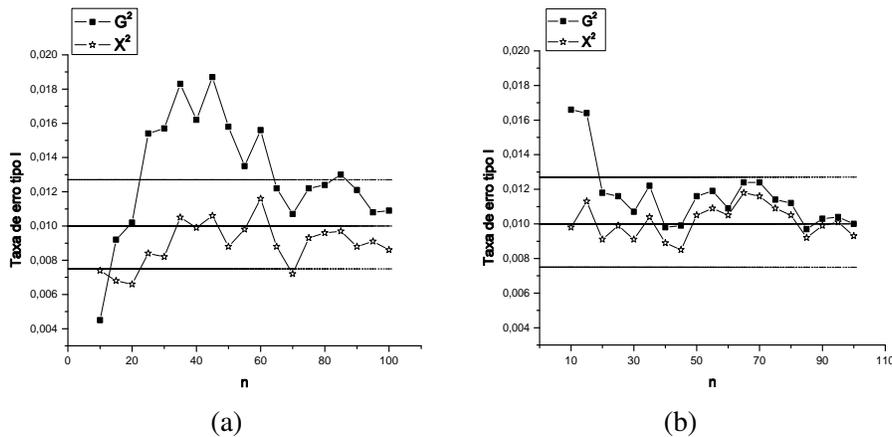


FIGURA 1.2 Taxas de erro tipo I dos testes de razão de verossimilhanças (G^2) e formas quadráticas (X^2), em função dos tamanhos amostrais (n) e valores dos parâmetros (a) $\pi = 0,1$ e (b) $\pi = 0,5$ para $\alpha = 1\%$, considerando a hipótese H_0 completa.

5.1.2 Poder sob H_0 parcial

Na Tabela 1.4 são apresentados os valores de poder dos testes G^2 e X^2 , em função de k , n e Δ , 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 π 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 Δ são pequenos ou moderados ($\Delta = 0,1$ e $0,3$), observa-se que o poder dos testes, em algumas situações, é pequeno (inferior a 70%), 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 se 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 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. 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 Δ , n e o teste. Porém, esse aumento foi sutilmente menor quando o tamanho das amostras e valores de Δ foi grande ($n \geq 30$ e $\Delta \geq 0,3$).

Para diferenças grandes ou muito grandes ($\Delta \geq 0,4$), verifica-se que o poder de ambos os testes aproxima-se de 100%, mesmo para pequenas amostras ($n = 10$). Assim, se o pesquisador tem interesse em averiguar pequenas diferenças ($\Delta < 0,01$) entre populações de grupos diferentes, é aconselhável estabelecer tamanhos de amostras

TABELA 1.4 Poder (%), sob H_0 parcial, para diferentes números de populações (k), de tamanhos de amostras (n) e de diferenças entre as proporções binomiais (Δ), para os testes de razão verossimilhanças (G^2) e formas quadráticas (X^2), ao nível nominal de 5%.

k	n	$\Delta = 0,01$		$\Delta = 0,1$		$\Delta = 0,3$	
		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
5	30	1,90	1,54	64,61	50,36	99,95	99,93
5	100	9,05	6,86	99,54	99,48	100,00	100,00
10	10	0,05	3,20	13,18	16,82	95,29	86,31
10	30	1,22	6,24	86,58	74,03	100,00	100,00
10	100	15,01	10,24	100,00	100,00	100,00	100,00
k	n	$\Delta = 0,4$		$\Delta = 0,5$		$\Delta = 0,6$	
		G^2	X^2	G^2	X^2	G^2	X^2
5	10	94,39	88,46	98,92	97,52	99,91	99,80
5	30	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
10	10	99,82	98,62	99,99	99,98	100,00	100,00
10	30	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
k	n	$\Delta = 0,7$		$\Delta = 0,8$		$\Delta = 0,9$	
		G^2	X^2	G^2	X^2	G^2	X^2
5	10	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
5	100	100,00	100,00	100,00	100,00	100,00	100,00
10	10	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
10	100	100,00	100,00	100,00	100,00	100,00	100,00

maiores que 100, principalmente se o número de populações for menor.

Na Tabela 1.5 são apresentados os valores de poder de ambos os testes em função de k , n e Δ , para $\alpha = 1\%$. O desempenho do poder dos testes, nesse caso, apresentou, de modo geral, semelhanças ao observado para o nível nominal de 5%. Entretanto, verifica-se que os valores de poder são maiores para $\alpha = 5\%$ do que para $\alpha = 1\%$, principalmente se as amostras são pequenas e intermediárias ($n \leq 30$) e se as diferenças Δ são pequenas ou moderadas ($\Delta \leq 0,3$), confirmando a validade da simulação.

TABELA 1.5 Poder (%), sob H_0 parcial, para diferentes números de populações (k), de tamanhos de amostras (n) e de diferenças entre as proporções binomiais (Δ), para os testes de razão verossimilhanças (G^2) e formas quadráticas (X^2), ao nível nominal de 1%.

k	n	$\Delta = 0,01$		$\Delta = 0,1$		$\Delta = 0,3$	
		G^2	X^2	G^2	X^2	G^2	X^2
5	10	0,01	0,02	1,32	2,39	43,74	35,40
5	30	0,17	0,14	34,52	26,37	99,67	99,37
5	100	1,98	1,76	97,94	97,40	100,00	100,00
10	10	0,01	0,33	2,69	5,60	79,25	60,06
10	30	0,10	0,77	62,30	44,10	100,00	100,00
10	100	3,70	3,04	99,99	99,97	100,00	100,00
k	n	$\Delta = 0,4$		$\Delta = 0,5$		$\Delta = 0,6$	
		G^2	X^2	G^2	X^2	G^2	X^2
5	10	76,35	65,54	93,90	88,47	99,14	97,91
5	30	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
10	10	97,80	91,12	99,94	99,33	100,00	100,00
10	30	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
k	n	$\Delta = 0,7$		$\Delta = 0,8$		$\Delta = 0,9$	
		G^2	X^2	G^2	X^2	G^2	X^2
5	10	99,97	99,74	100,00	100,00	100,00	100,00
5	30	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
10	10	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
10	100	100,00	100,00	100,00	100,00	100,00	100,00

Pode-se observar que, para $\Delta = 0,01$, os valores de poder dos dois testes avaliados são pequenos, para todos os tamanhos amostrais e número de populações. Para $\Delta = 0,1$, o poder dos testes foi consideravelmente pequeno (inferior a 70%) para tamanhos amostrais iguais a 10 e 30, tanto para $k = 5$ quanto para $k = 10$. Considerando $\Delta = 0,3$, verificou-se que o poder de ambos os testes é pequeno para pequenas amostras ($n = 10$) e $k = 5$ para ambos os testes e também com $k = 10$, $n = 10$ para o teste X^2 . Com o aumento dos valores de n houve um crescimento expressivo do poder dos testes,

atingindo 100% para grandes amostras ($n = 100$). É conveniente enfatizar que, para as situações em que $\pi = 0,01$, sob H_0 completa, houve controle do erro tipo I, tanto para $\alpha = 5\%$ quanto para $\alpha = 1\%$ (Tabelas 1.2 e 1.3). Esse controle ocorreu, de maneira geral, de forma conservativa.

Se os valores de Δ são grandes ou muito grandes ($\Delta \geq 0,4$), conforme já salientado para $\alpha = 5\%$, o poder de ambos os testes tende a igualar e se aproximar de 100%. No entanto, observa-se que, o poder do teste X^2 é relativamente pequeno para pequenas amostras ($n = 10$), considerando $\Delta = 0,4$ e $k = 5$.

O desempenho do poder dos testes estudados também foi avaliado em algumas situações em que os valores de π 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 π 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 Δ , ou seja, $\pi^{(2)} = \pi^{(1)} + \Delta$. Nessas situações específicas, os valores de Δ foram iguais a 0,01, 0,1 e 0,4.

Na Tabela 1.6 são apresentados os valores de poder dos testes G^2 e X^2 em função de k , $\pi^{(1)}$, n e Δ , 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 (inferiores a 70%), 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 $n = 10$, principalmente em relação ao teste X^2 . Para amostras intermediárias e grandes ($n \geq 30$), os valores de poder de ambos os testes tenderam a se igualar e se aproximaram de 100%.

TABELA 1.6 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 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

Observou-se que, se a diferença entre os valores de π 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 1.6, fixando-se 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 desempenho também foi semelhante ao observado para valores de π afastado de 0,5 (Tabela 1.4).

Para o nível nominal de significância de 1%, os valores de poder de ambos os testes simulados apresentaram grandes similaridades aos resultados obtidos para $\alpha =$

5% e são apresentados na Tabela 1.7. Em função disso, vale ressaltar, como já esperado, que o poder de ambos os testes é relativamente inferior aos observados para o nível nominal de 5%, principalmente se os valores de Δ são pequenos ou moderados ($\Delta \leq 0,1$).

TABELA 1.7 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 os testes de razão verossimilhanças (G^2) e formas quadráticas (X^2), ao nível nominal de 1%.

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	2,19	0,95	2,95	1,36	40,21	33,88
5	0,30	30	1,27	0,93	4,81	4,22	95,28	94,76
5	0,30	100	1,05	0,97	20,37	20,16	100,00	100,00
10	0,30	10	2,12	0,84	3,45	1,57	66,72	58,25
10	0,30	30	1,38	1,00	6,70	5,59	99,90	99,89
10	0,30	100	1,44	1,41	38,58	37,62	100,00	100,00
5	0,45	10	1,81	1,04	2,65	1,53	46,59	34,23
5	0,45	30	1,07	0,86	4,12	3,61	97,41	96,72
5	0,45	100	0,96	0,94	18,18	17,80	100,00	100,00
10	0,45	10	1,90	0,83	3,06	1,33	76,21	65,40
10	0,45	30	1,06	0,84	5,74	4,80	99,98	99,97
10	0,45	100	0,89	0,82	32,95	32,24	100,00	100,00
5	0,50	10	1,56	0,87	2,47	1,34	53,53	37,07
5	0,50	30	1,19	0,97	4,10	3,59	98,87	98,40
5	0,50	100	1,14	1,07	18,97	18,42	100,00	100,00
10	0,50	10	1,89	0,81	3,14	1,46	83,08	71,98
10	0,50	30	1,22	0,95	6,20	5,12	100,00	100,00
10	0,50	100	1,23	1,15	33,35	32,54	100,00	100,00

Se as situações de $\pi^{(1)} = 0,01$ e $\pi^{(1)} = 0,30, 0,45$ e $0,50$ (Tabelas 1.4 e 1.6) ou (Tabelas 1.5 e 1.7) forem comparadas, pode-se observar que o poder aumenta com o aumento da diferença paramétrica Δ 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 $0,5$ (Tabelas 1.6 e 1.7), o aumento do valor de poder com o aumento de Δ

é menor do que esse mesmo aumento com $\pi^{(1)} = 0,01$ (Tabelas 1.4 e 1.5). 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 Δ , o poder é maior na medida que $\pi^{(1)}$ aumenta e se aproxima de 0,5.

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 é apresentado na Figura 1.3, em função de Δ com $k = 5$ e $n = 10$, considerando $\alpha = 5$ e 1%. Para $\alpha = 5\%$ (Figura 1.3 a e c), 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 Δ , os valores de poder dos testes para $\pi^{(1)} = 0,01$ foram relativamente superiores, se comparados 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.

Para $\alpha = 1\%$ (Figura 1.3 b e d), observou-se o mesmo desempenho geral dos testes. No entanto, verificou-se, para $\Delta \leq 0,1$, que os valores de poder de ambos os testes foram aproximadamente iguais ao valor nominal, independente dos valores de $\pi^{(1)}$. Para $\Delta = 0,4$, as diferenças dos valores de poder entre $\pi^{(1)} = 0,01$ e $\pi^{(1)} = 0,30, 0,45$ e $0,50$ (Figura 1.3 d) do teste X^2 foram expressivamente maiores que as observadas para o teste G^2 (Figura 1.3 b). Ainda para $\Delta = 0,4$, observa-se que os valores de poder do teste X^2 , para todos os valores de $\pi^{(1)}$ próximos de 0,5 ($\pi^{(1)} = 0,30, 0,45, 0,50$), foram pequenos e muito semelhantes.

A superioridade do teste G^2 em relação ao teste X^2 , considerando o poder, tem de ser vista com certa ressalva, devido ao fato de as taxas de erro tipo I do teste G^2 serem superiores aos níveis nominais em pequenas amostras ($n = 10$) e valores de π

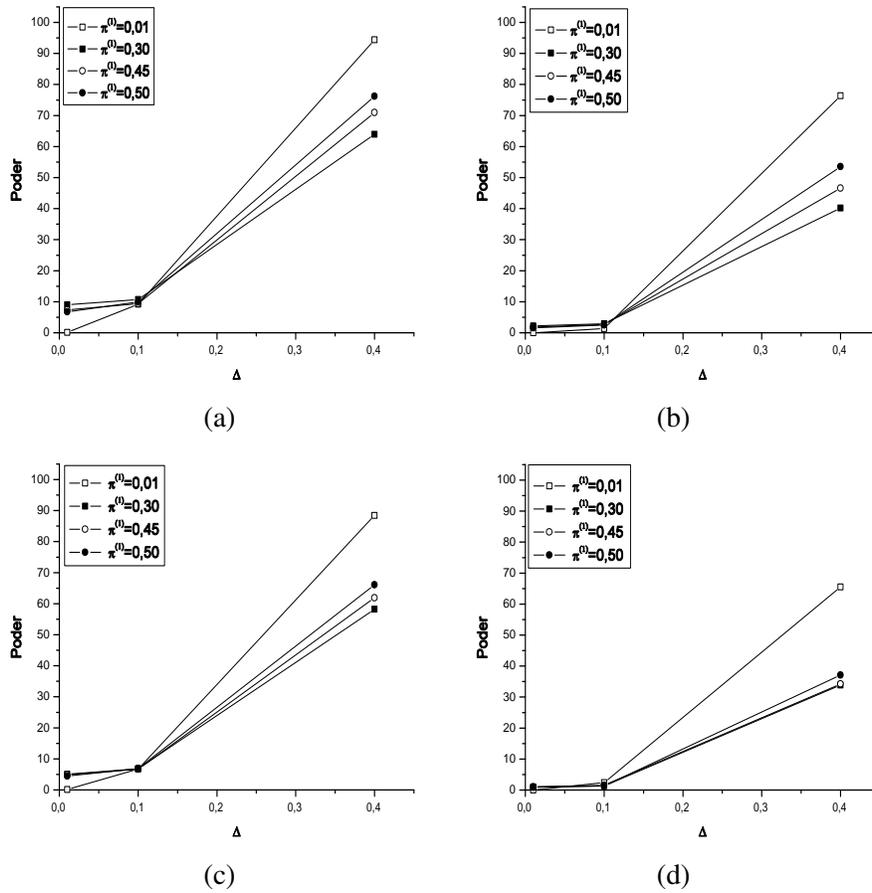


FIGURA 1.3 Poder, sob H_0 parcial, dos testes G^2 e X^2 , em função da diferença Δ , com $k = 5$, $n = 10$ e valores nominais de significância iguais (a) $\alpha = 5\%$ e (b) $\alpha = 1\%$ para o teste G^2 e (c) $\alpha = 5\%$ e (d) $\alpha = 1\%$, para o teste X^2 .

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 α , 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 por experimento.

5.2 Comparações múltiplas

A estatística (4.11) para testar a hipótese (4.10), utilizando-se o estimador da variância, é simplificada por:

$$\chi_{c3}^2 = \frac{(\hat{\pi}_i - \hat{\pi}_{i'})^2}{\frac{\hat{\pi}_i(1-\hat{\pi}_i)}{n_i} + \frac{\hat{\pi}_{i'}(1-\hat{\pi}_{i'})}{n_{i'}}}, \quad (5.1)$$

uma vez que $\text{var}(\hat{\pi}_i) = \hat{\pi}_i(1 - \hat{\pi}_i)/n_i$, $\text{var}(\hat{\pi}_{i'}) = \hat{\pi}_{i'}(1 - \hat{\pi}_{i'})/n_{i'}$ e a $\text{cov}(\hat{\pi}_i, \hat{\pi}_{i'}) = 0$. Sob a hipótese nula, χ_{c3}^2 possui distribuição assintótica qui-quadrado com $\nu = k - 1$ graus de liberdade. Se $\chi_{c3}^2 \geq \chi_{\alpha, k}^2$, a hipótese (4.10) deve ser rejeitada.

No entanto, essa expressão possui muitas limitações práticas, principalmente se n_i for pequeno ($n \leq 10$) e se os valores de π_i se aproximam de 0 ou de 1. A razão disso é que podem ocorrer denominadores nulos na expressão (5.1) em decorrência de estimativas nulas das variâncias. Assim, para contornar essa limitação, foi proposto o estimador comum (4.5) dos π_i para ser utilizado no estimador das variâncias dos $\hat{\pi}_i$. Assim, a estatística do teste foi dada por:

$$\chi_{c4}^2 = \frac{(\hat{\pi}_i - \hat{\pi}_{i'})^2}{\frac{\hat{\pi}_0(1-\hat{\pi}_0)}{n_i} + \frac{\hat{\pi}_0(1-\hat{\pi}_0)}{n_{i'}}}, \quad (5.2)$$

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

Nesse caso, as estimativas de π_0 podem ser nulas ou iguais à unidade em algumas situações, o que levaria o denominador de (5.2) 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 χ_{c4}^2 como nulo. Da mesma forma, χ_{c4}^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. Essas duas situações são discutidas nas subseções seguintes, 5.2.1 e 5.2.2. Em seguida, o poder do TCM foi avaliado sob a hipótese H_0 parcial.

5.2.1 Erro tipo I sob H_0 completa

Na Tabela 1.8 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 π , considerando o nível nominal de 5%. O que se observa, de maneira geral, é que, na grande maioria dos casos, houve controle do erro tipo I por experimento, que ocorreu de forma conservativa. Em apenas um caso, o TCM apresentou taxas de erro tipo I significativamente superiores ao nível nominal de 5% e, em outros casos, as taxas de erro tipo I do TCM foram não significativamente ($P > 0,01$) diferentes do valor nominal. Pode-se verificar também que houve excesso de conservadorismo, para os casos em que os números de populações eram grandes ($k = 5$ e 10), independentemente dos tamanhos amostrais e dos valores de π .

O TCM foi considerado liberal com $k = 2$, $n = 100$ para $\pi = 0,5$. No entanto, verificou-se que essa diferença não é expressivamente superior ao valor nominal adotado de 5%. O TCM foi considerado exato nas situações com $k = 2$, $n = 30$ e 100 para $\pi = 0,1$ e com $k = 2$, $n = 30$ para $\pi = 0,5$. Em todos os demais casos, o TCM apresentou taxas de erro tipo I significativamente ($P < 0,01$) inferiores ao valor nominal de 5%.

Biase (2006) avaliou as taxas de erro tipo I por experimento em situações similares às consideradas neste trabalho sob H_0 completa, para dois 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 $\pi = 0,1$, $0,5$ e $0,9$, com $\alpha = 5\%$, esses

TABELA 1.8 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 π para o teste de comparações múltiplas (TCM), ao nível nominal de 5%.

k	n	$\pi = 0,01$	$\pi = 0,1$	$\pi = 0,5$
2	10	0,00 ⁺	0,92 ⁺	4,14 ⁺
2	30	0,04 ⁺	5,30 ^{ns}	5,30 ^{ns}
2	100	1,23 ⁺	5,09 ^{ns}	5,61 [*]
5	10	0,00 ⁺	0,31 ⁺	2,14 ⁺
5	30	0,03 ⁺	1,30 ⁺	2,10 ⁺
5	100	0,32 ⁺	1,45 ⁺	1,86 ⁺
10	10	0,00 ⁺	0,06 ⁺	0,04 ⁺
10	30	0,00 ⁺	0,10 ⁺	0,20 ⁺
10	100	0,13 ⁺	0,10 ⁺	0,09 ⁺

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

⁺ significativamente inferior ao nível nominal de 5%, considerando uma confiança de 99%.

^{ns} não significativamente diferente do nível nominal de 5%, considerando uma confiança de 99%.

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 amostras pequenas e intermediárias ($n = 10$ ou 30) e valores de π afastados de $0,5$. Para grandes amostras ($n = 100$), os dois testes foram considerados exatos, exceto com maior número de populações ($k = 10$) e $\pi = 0,9$, em que o teste de *bootstrap* Pan foi considerado conservativo.

O TCM apresentou algumas diferenças em relação ao desempenho dos testes de *bootstrap* sendo considerado conservativo para todas as situações em que o número de populações é elevado ($k = 5$ e 10) e liberal em uma situação com pequeno número de populações ($k = 2$), conforme mencionado anteriormente. Isso não ocorreu com os testes de *bootstrap*. No entanto, é importante salientar que os testes de *bootstrap* exigem maior esforço computacional, se comparados com o TCM avaliado neste 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, têm 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 & 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*.

Para o nível nominal de significância de 1%, as taxas de erro tipo I, por experimento, do TCM são apresentadas na Tabela 1.9. O desempenho do TCM teve inúmeras similaridades com os resultados observados para 5%. Pode-se observar que para grandes amostras ($n = 100$), com $k = 2$ e $\pi = 0,1$ e $0,5$ e também com $k = 2$, $n = 10$ e $\pi = 0,5$, o TCM apresentou tamanho não significativamente ($P > 0,01$) diferente do valor nominal de significância de 1%. O TCM foi liberal apenas para $k = 2$ com $n = 30$ e $\pi = 0,5$. Nas demais situações, o TCM foi conservativo, ou seja, apresentou taxas de erro tipo I significativamente ($P < 0,01$) inferiores ao valor nominal.

De modo geral, o TCM foi mais conservativo considerando $\alpha = 1\%$, para $k = 5$ e 10 . Observou-se também que os tamanhos amostrais não influenciaram as taxas de erro tipo I por experimento, pois a variação destas taxas, na maioria dos casos, foi muito pequena. Tanto para $\alpha = 5\%$ como para $\alpha = 1\%$, verifica-se que, houve uma redução das taxas com o aumento de k de 2 para 5. De 5 para 10, em alguns casos, essa redução foi ligeiramente menor, considerando $\pi = 0,1, 0,5$.

TABELA 1.9 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 π para o teste de comparações múltiplas (TCM), ao nível nominal de 1%.

k	n	$\pi = 0,01$	$\pi = 0,1$	$\pi = 0,5$
2	10	0,00 ⁺	0,13 ⁺	1,22 ^{ns}
2	30	0,00 ⁺	0,68 ⁺	1,34 [*]
2	100	0,00 ⁺	0,89 ^{ns}	0,82 ^{ns}
5	10	0,00 ⁺	0,03 ⁺	0,07 ⁺
5	30	0,00 ⁺	0,17 ⁺	0,12 ⁺
5	100	0,00 ⁺	0,23 ⁺	0,16 ⁺
10	10	0,00 ⁺	0,00 ⁺	0,00 ⁺
10	30	0,00 ⁺	0,01 ⁺	0,02 ⁺
10	100	0,02 ⁺	0,01 ⁺	0,01 ⁺

* significativamente superior ao nível nominal de 1% considerando uma confiança de 99%.

⁺ significativamente inferior ao nível nominal de 1% considerando uma confiança de 99%.

^{ns} não significativamente diferente do nível nominal de 1%, considerando uma confiança de 99%.

5.2.2 Erro tipo I sob H_0 parcial

Na Tabela 1.10, as taxas de erro tipo I por experimento sob H_0 parcial para o TCM são apresentadas em função de k , n e Δ , 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 π 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 π 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 π de ambos os grupos extrapolassem o espaço paramétrico dos parâmetros binomiais π , que variam entre 0 e 1.

Comparando-se as taxas de erro tipo I para $k = 5$ e $k = 10$, observa-se, para um valor fixo de Δ , que houve redução das taxas quando o número de populações aumentou de 5 para 10, independentemente dos tamanhos amostrais. Esse desempenho foi

TABELA 1.10 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 (TCM), ao nível nominal de 5%.

k	n	Δ					
		0,01	0,05	0,1	0,2	0,3	0,4
5	10	0,00 ⁺	0,18 ⁺	0,80 ⁺	2,52 ⁺	2,66 ⁺	1,91 ⁺
5	30	0,14 ⁺	1,93 ⁺	2,81 ⁺	3,17 ⁺	2,65 ⁺	1,82 ⁺
5	100	0,71 ⁺	2,30 ⁺	3,13 ⁺	3,22 ⁺	2,61 ⁺	1,81 ⁺
10	10	0,00 ⁺	0,08 ⁺	0,49 ⁺	0,89 ⁺	0,66 ⁺	0,38 ⁺
10	30	0,09 ⁺	0,78 ⁺	1,14 ⁺	1,15 ⁺	0,90 ⁺	0,58 ⁺
10	100	0,16 ⁺	1,07 ⁺	1,24 ⁺	1,22 ⁺	0,99 ⁺	0,61 ⁺

k	n	Δ				
		0,5	0,6	0,7	0,8	0,9
5	10	1,35 ⁺	0,57 ⁺	0,27 ⁺	0,03 ⁺	0,00 ⁺
5	30	1,39 ⁺	0,64 ⁺	0,26 ⁺	0,04 ⁺	0,00 ⁺
5	100	1,19 ⁺	0,69 ⁺	0,22 ⁺	0,03 ⁺	0,00 ⁺
10	10	0,26 ⁺	0,08 ⁺	0,02 ⁺	0,00 ⁺	0,00 ⁺
10	30	0,38 ⁺	0,10 ⁺	0,02 ⁺	0,00 ⁺	0,00 ⁺
10	100	0,24 ⁺	0,07 ⁺	0,01 ⁺	0,00 ⁺	0,00 ⁺

⁺ significativamente inferior ao nível nominal de 5%, considerando uma confiança de 99%.

semelhante ao observado para as taxas de erro tipo I por experimento, sob H_0 completa.

Para o nível de significância de 1%, as taxas de erro tipo I por experimento do TCM são apresentadas na Tabela 1.11, em função de k , n e Δ . Pode-se observar, de modo geral, que o TCM controlou as taxas de erro tipo I em nível inferior ou, no máximo, igual ao valor nominal de 1%. Os casos em que o TCM teve taxas de erro tipo I não significativamente ($P > 0,01$) diferentes de $\alpha = 1\%$ foram com $k = 5$, $n = 10$ para $\Delta = 0,2$ e $0,3$ e com $k = 5$, $n = 30$ e 100 considerando $\Delta = 0,1$, $0,2$ e $0,3$.

De maneira análoga ao que ocorreu para $\alpha = 5\%$, as taxas de erro tipo I foram expressivamente inferiores ao valor nominal de 1%, para valores de π afastados de 0,5. Isso ocorreu para todos os tamanhos amostrais. Esse fato pode ter como consequência pequenos valores de poder, o que intensifica o risco de não detectar diferenças significativas entre os tratamentos, se Δ for pequeno.

TABELA 1.11 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 (TCM), ao nível nominal de 1%.

k	n	Δ					
		0,01	0,05	0,1	0,2	0,3	0,4
5	10	0,00 ⁺	0,02 ⁺	0,10 ⁺	0,85 ^{ns}	1,05 ^{ns}	0,65 ⁺
5	30	0,00 ⁺	0,19 ⁺	0,75 ^{ns}	0,82 ^{ns}	0,88 ^{ns}	0,53 ⁺
5	100	0,08 ⁺	0,67 ⁺	0,89 ^{ns}	1,10 ^{ns}	0,86 ^{ns}	0,56 ⁺
10	10	0,00 ⁺	0,02 ⁺	0,10 ⁺	0,16 ⁺	0,09 ⁺	0,05 ⁺
10	30	0,01 ⁺	0,12 ⁺	0,29 ⁺	0,21 ⁺	0,24 ⁺	0,12 ⁺
10	100	0,04 ⁺	0,17 ⁺	0,29 ⁺	0,38 ⁺	0,20 ⁺	0,14 ⁺

k	n	Δ				
		0,5	0,6	0,7	0,8	0,9
5	10	0,50 ⁺	0,19 ⁺	0,04 ⁺	0,00 ⁺	0,00 ⁺
5	30	0,32 ⁺	0,13 ⁺	0,04 ⁺	0,00 ⁺	0,00 ⁺
5	100	0,28 ⁺	0,11 ⁺	0,04 ⁺	0,00 ⁺	0,00 ⁺
10	10	0,02 ⁺	0,00 ⁺	0,00 ⁺	0,00 ⁺	0,00 ⁺
10	30	0,04 ⁺	0,00 ⁺	0,00 ⁺	0,00 ⁺	0,00 ⁺
10	100	0,06 ⁺	0,00 ⁺	0,00 ⁺	0,00 ⁺	0,00 ⁺

⁺ significativamente inferior ao nível nominal de 1%, considerando uma confiança de 99%.

^{ns} não significativamente diferente do nível nominal de 1%, considerando uma confiança de 99%.

Os resultados das taxas de erro tipo I por experimento do TCM, tanto para $\alpha = 5\%$ quanto para $\alpha = 1\%$, são coerentes com os resultados obtidos para os testes de *bootstrap* avaliados por Biase (2006), que também apresentaram, na 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 controlaram o erro tipo I sob a hipótese H_0 parcial e apresentaram desempenho bastante similar para todos os tamanhos de amostras n , número de populações k e valores de Δ .

As taxas de erro tipo I por experimento, sob a hipótese H_0 parcial, também foram avaliadas em algumas situações em que os valores de π dos dois grupos se apro-

ximavam de 0,5 tanto quanto fosse possível. Para tal finalidade, foram realizadas simulações adicionais, considerando os níveis nominais de significância de 1% e 5%. Como as taxas de erro tipo I por experimento do TCM, ao nível nominal de 1%, foram semelhantes as taxas considerando $\alpha = 5\%$, serão apresentados apenas os resultados obtidos para $\alpha = 5\%$. Na Tabela 1.12 são apresentadas as taxas de erro tipo I por experimento, sob H_0 parcial, do TCM em função de k , $\pi^{(1)}$, n e Δ para $\alpha = 5\%$.

Os resultados foram todos conservativos, embora sejam menos conservativos se comparados aos observados sob H_0 parcial, com um dos grupos afastados grandemente de 0,5. O mesmo desempenho do TCM para valores de π próximos de 0,5 foi observado em relação ao efeito do número de populações, no sentido de reduzir os valores das taxas de erro tipo I. Assim, observa-se que as taxas de erro tipo I diminuem com o aumento de $k = 5$ para $k = 10$, se fixado $\pi^{(1)}$, n e Δ . Outro aspecto que deve ser abordado é que, com o aumento do tamanho das amostras, as taxas de erro tipo I mantiveram o padrão conservativo, para todas as configurações de k , $\pi^{(1)}$ e Δ . Para os testes de *bootstrap* apresentados por Biase (2006), observou-se esse mesmo desempenho dos testes, considerando valores de π dos grupos próximos de 0,5.

TABELA 1.12 Taxas de erro tipo I por experimento (%), 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 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	0,59 ⁺	0,63 ⁺	0,33 ⁺
5	0,30	30	0,71 ⁺	0,52 ⁺	0,49 ⁺
5	0,30	100	0,77 ⁺	0,68 ⁺	0,36 ⁺
10	0,30	10	0,01 ⁺	0,02 ⁺	0,00 ⁺
10	0,30	30	0,04 ⁺	0,04 ⁺	0,00 ⁺
10	0,30	100	0,07 ⁺	0,11 ⁺	0,01 ⁺
5	0,45	10	0,87 ⁺	0,84 ⁺	0,82 ⁺
5	0,45	30	1,03 ⁺	0,85 ⁺	0,72 ⁺
5	0,45	100	0,75 ⁺	1,03 ⁺	0,62 ⁺
10	0,45	10	0,00 ⁺	0,01 ⁺	0,01 ⁺
10	0,45	30	0,07 ⁺	0,06 ⁺	0,03 ⁺
10	0,45	100	0,08 ⁺	0,03 ⁺	0,06 ⁺
5	0,50	10	1,02 ⁺	0,96 ⁺	0,95 ⁺
5	0,50	30	1,12 ⁺	0,97 ⁺	0,85 ⁺
5	0,50	100	1,07 ⁺	0,76 ⁺	1,06 ⁺
10	0,50	10	0,01 ⁺	0,02 ⁺	0,06 ⁺
10	0,50	30	0,10 ⁺	0,06 ⁺	0,23 ⁺
10	0,50	100	0,06 ⁺	0,03 ⁺	0,19 ⁺

⁺ significativamente inferior ao nível nominal de 5%, considerando uma confiança de 99%.

5.2.3 Poder sob H_0 parcial

Na Tabela 1.13 são apresentados os valores de poder, em porcentagem, para o TCM, em função de k , n e Δ para $\alpha = 5\%$. Para valores muito pequenos de Δ ($\Delta = 0,01$ e $0,05$), observa-se que os valores de poder são extremamente pequenos (inferiores a 70%), independente do número de populações e do tamanho das amostras. Para valores pequenos e moderados de Δ ($0,1 \leq \Delta \leq 0,3$), 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 1.13 Poder (%), sob H_0 parcial, para diferentes números de populações (k), tamanhos de amostras (n) e diferenças entre os parâmetros π de cada grupo (Δ), para o teste de comparações múltiplas (TCM), ao nível nominal de 5%.

k	n	Δ					
		0,01	0,05	0,1	0,2	0,3	0,4
5	10	0,00	0,09	0,48	3,25	8,50	18,59
5	30	0,07	1,53	7,46	34,51	67,72	89,83
5	100	0,53	14,83	56,63	97,99	99,99	100,00
10	10	0,00	0,02	0,10	0,24	0,40	0,94
10	30	0,02	0,16	0,57	3,74	13,89	35,34
10	100	0,03	1,25	10,96	67,30	97,51	99,97

k	n	Δ				
		0,5	0,6	0,7	0,8	0,9
5	10	33,77	51,10	69,69	88,35	98,62
5	30	98,09	99,88	100,00	100,00	100,00
5	100	100,00	100,00	100,00	100,00	100,00
10	10	1,91	4,91	11,79	15,67	35,03
10	30	65,37	89,55	98,37	99,95	100,00
10	100	100,00	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 se aproximaram de 100%. Considerando $\Delta = 0,5$ e $0,6$, o poder do TCM foi pequeno 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 (superiores a 70%) e tenderam a se aproximar de 100%, exceto para $\Delta = 0,5$ com $k = 10$ e $n = 30$. Se as diferenças (Δ) são muito grandes ($\Delta \geq 0,7$), os valores de poder do TCM são altos e tendem a se aproximar de 100%, exceto para $k = 10$ e $n = 10$ e $k = 5$, $n = 10$, considerando $\Delta = 0,7$.

De maneira geral, verifica-se, que para todos os valores de Δ , houve 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.

Para o nível nominal de 1%, o desempenho geral do TCM foi semelhante aos obtidos para $\alpha = 5\%$. Na Tabela 1.14 são apresentados os valores de poder do TCM em função de k , n e Δ , para $\alpha = 1\%$. De modo geral, pode-se observar que, na maioria das situações avaliadas, considerando pequenas amostras ($n = 10$), o poder do TCM foi extremamente pequeno (inferior a 70%), principalmente para $k = 10$. A exceção ocorreu para $k = 5$ e $\Delta = 0,9$. Da mesma forma que ocorreu para $\alpha = 5\%$, verificou-se que os valores de poder do TCM tiveram crescimento com o aumento do tamanho das amostras, aproximando-se de 100% em muitos casos e, ainda, observou-se redução do poder do TCM com o aumento de k .

TABELA 1.14 Poder (%), sob H_0 parcial, para diferentes números de populações (k), tamanhos de amostras (n) e diferenças entre os parâmetros π de cada grupo (Δ), para o teste de comparações múltiplas (TCM), ao nível nominal de 1%.

k	n	Δ					
		0,01	0,05	0,1	0,2	0,3	0,4
5	10	0,00	0,01	0,06	0,88	2,37	6,08
5	30	0,00	0,21	1,85	14,59	42,84	74,06
5	100	0,05	4,62	32,65	92,52	99,87	100,00
10	10	0,00	0,00	0,02	0,04	0,05	0,10
10	30	0,00	0,02	0,10	0,92	4,45	14,87
10	100	0,01	0,29	3,72	43,25	90,61	99,70

k	n	Δ				
		0,5	0,6	0,7	0,8	0,9
5	10	12,88	23,58	41,30	68,07	92,17
5	30	92,48	99,01	99,95	100,00	100,00
5	100	100,00	100,00	100,00	100,00	100,00
10	10	0,15	0,49	0,71	0,10	0,00
10	30	37,98	70,56	92,21	99,62	100,00
10	100	100,00	100,00	100,00	100,00	100,00

Para $\Delta \leq 0,1$, os valores de poder foram pequenos para todas as configurações de k e n . Se os valores de Δ são moderados ou grandes ($\Delta = 0,2, 0,3$ e $0,4$), o poder do TCM é alto para grandes amostras ($n = 100$), exceto para $k = 10, n = 100$ e $\Delta = 0,2$.

Para diferenças muito grandes de Δ ($\Delta \geq 0,5$), os valores de poder aproximam de 100% para $n \geq 30$, com exceção para $k = 10$, $n = 30$ com $\Delta = 0,5$.

Biase (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 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 dos TCM avaliado neste trabalho foi verificado com o aumento de n e k . O mesmo desempenho 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 & Ferreira, 2003; Santos et al., 2001)

Procurando avaliar situações em que os valores de π dos grupos se aproximavam de 0,5, foram realizadas algumas simulações adicionais para mensurar o poder do TCM sob a hipótese H_0 parcial, considerando os níveis nominais de 1% e 5%. Novamente, verificou-se que houve uma semelhança muito grande entre os resultados de 1% e 5% e, por isso, são apresentados apenas os resultados obtidos para $\alpha = 5\%$.

Na Tabela 1.15 são apresentados os valores de poder do TCM, em função de k , $\pi^{(1)}$, n e Δ , para $\alpha = 5\%$. Pode-se observar, para $\Delta = 0,01$ e $0,1$, que o poder do TCM foi expressivamente pequeno e que, em todas as situações considerando $k = 10$, o poder foi inferior ou próximo do valor nominal de 5%. Para uma diferença maior entre os grupos ($\Delta = 0,4$), os valores de poder foram pequenos (inferiores a 70%) com $n \leq 30$ e, para grandes amostras ($n = 100$), os valores de poder do TCM se aproximaram de 100%. Esses resultados estão de acordo com os apresentados por Biase (2006), considerando os testes de *bootstrap*.

De modo geral, verifica-se que, se os valores de π 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

TABELA 1.15 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 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

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 redução com o aumento de $k = 5$ para $k = 10$. Esse último fato já poderia ter sido antecipado quando se avaliou o erro tipo I, pois, com o aumento de k de 5 para 10, sob H_0 parcial, verificou-se que houve tendência de o TCM ser mais conservativo (Tabela 1.12).

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 do que os que são praticados normalmente sejam utilizados nas pesquisas, quando isso for viável de ser utilizado.

Em decorrência desse fato, foram realizadas algumas simulações extras para avaliar o TCM em relação às 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 π no primeiro grupo fossem iguais a $0,05$ e os do segundo, por $0,05 + \Delta$.

Na Tabela 1.16 são apresentadas as taxas de erro tipo I por experimento sob H_0 parcial em função de k , n e Δ , 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 Δ , sendo considerados conservativos.

Na Tabela 1.17 são apresentados os valores de poder do TCM em função de k , n e Δ , para $\alpha = 5\%$. Verifica-se, de modo geral, que os valores de poder foram pequenos em todas as situações, exceto para $\Delta = 0,04$ e $k = 5$, com $n = 800$ 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 Δ , o poder do TCM aumenta.

TABELA 1.16 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 π de cada grupo (Δ) 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 ⁺	0,54 ⁺	0,73 ⁺	0,99 ⁺
5	800	0,72 ⁺	0,79 ⁺	0,90 ⁺	1,01 ⁺
5	1000	0,69 ⁺	0,77 ⁺	0,85 ⁺	0,85 ⁺
10	400	0,07 ⁺	0,16 ⁺	0,17 ⁺	0,35 ⁺
10	800	0,09 ⁺	0,12 ⁺	0,23 ⁺	0,29 ⁺
10	1000	0,10 ⁺	0,06 ⁺	0,09 ⁺	0,29 ⁺

⁺ significativamente inferior ao nível nominal de 5%, considerando uma confiança de 99%.

TABELA 1.17 Poder (%), sob H_0 parcial, para diferentes números de populações (k), de tamanhos de amostras (n) e de diferenças entre os parâmetros π de cada grupo (Δ) 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

6 Conclusões

O teste X^2 controlou o erro tipo I em níveis iguais ou inferiores aos valores nominais de significância e apresentou desempenho superior ao do teste G^2 , que foi considerado liberal, principalmente nas situações de pequenas amostras e maiores números de populações.

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.

7 Referências Bibliográficas

- AGRESTI, A.; BINI, M.; BERTACCINI, B.; RYU, E. Simultaneous confidence intervals for comparing binomial parameters. **Biometrics**, Washington, v.68, n.4, p.1-6, Dec. 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, Dec. 1987.
- BIASE, N.G. **Comparações múltiplas para parâmetros binomiais utilizando bootstrap**. 2006. 73p. Dissertação (Mestrado em Agronomia) - Universidade Federal de Lavras, Lavras.
- BIASE, N.G.; FERREIRA, D.F. Comparações múltiplas para proporções binomiais utilizando bootstrap. **Revista de Matemática e Estatística**, 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. **Revista de Matemática e Estatística**, 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 and Hall, 2008, 244p.
- FERREIRA, D.F. **Estatística multivariada**. Lavras: UFLA, 2008. 662p. 1 v.
- HOCHBERG, Y.; TAMHANE, A.C. **Multiple comparison procedures**. New York: J. Wiley, 1987. 450p.
- KIM, B.S.; ZHAO, B.; KIM, H.J.; CHO, M. H. The statistical analysis of the in vitro chromosome aberration assay using Chinese hamster ovary cells. **Mutation research-genetic toxicology and environmental mutagenesis**, v.469, p.243-252, June 2000.
- KRISHNAMOORTHY, K.; PENG, J. Exact properties of a new test and other tests for differences between several binomial proportions. **Journal of Applied Statistical Science**, 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. **Computational Statistics & Data Analysis**, Amsterdam, v.45, p.697-707, 2004.
- MCCULLOCH, C.E.; SEARLE, S.R. **Generalized, linear, and mixed models**. New York: J. Wiley, 2001. 346p.
- MOOD, A.M.; GRAYBILL, F.A.; BOES, D.C. **Introduction to the theory of statistics**. 3. ed. New York: J. Wiley, 1974. 564p.
- NELDER, J.A.; WEDDERBURN, R.W.M. Generalized linear models. **Journal of the Royal Statistical Society**, London, v.135, p.370-384, 1972.
- PAN, W. Approximate confidence intervals for one proportion and two proportions. **Computational Statistics & Data Analysis**, Amsterdam, v.40, n.1, p.143-157, July 2002.
- PIEGORSCH, W.W. Multiple comparisons for analysing dichotomous response. **Biometrics**, Washington, v.47, n.1, p.45-52, Mar. 1991.
- RAVEN, P.H.; EVERT, R.F.; EICHHOR, S.E. **Biology of plants**. 7. ed. New York: Freeman and Company, 2005. 960p.
- R Development Core Team. **R: A language and environment for statistical computing**. Vienna, Austria: R Foundation for Statistical Computing, 2008. Disponível em: <<http://www.r-project.org>>. Acesso em: 10 nov. 2008.
- SANTOS, C.; FERREIRA, D.F.; BUENO FILHO, J.S.S. Novas alternativas de testes de agrupamento avaliadas por meio de simulação Monte Carlo. **Ciência e Agrotecnologia**, Lavras, v.25, n.6, p.1382-1392, nov. 2001.
- WILLIAMS, D.A. Test for differences between several small proportions. **Journal of the Royal Statistical Society**, London, v.37, n.3, p.421-434, 1988.

CAPÍTULO 3

INFERÊNCIA BAYESIANA PARA k POPULAÇÕES BINOMIAIS INDEPENDENTES

1 RESUMO

Na teoria frequentista, inferências sobre duas ou mais 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 de algumas das pressuposições exigidas pelos testes ou, então, à necessidade de implementação de rotinas para suas aplicações. Por essa razão, este trabalho teve por objetivos propor uma abordagem bayesiana para realizar um teste de comparações múltiplas de proporções binomiais e um teste global de igualdade de várias proporções binomiais e, ainda, avaliar o desempenho desses testes utilizando simulação Monte Carlo. Foram geradas k populações binomiais independentes com parâmetros π_i e n_i , $i = 1, 2, \dots, k$ e realizadas 1.000 simulações 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. As simulações foram subdivididas em duas partes, sendo a primeira para o teste bayesiano (TB) e a segunda para o teste de comparações múltiplas bayesianas (TCMB). Cada uma delas foi subdividida novamente em duas etapas. Na primeira, foram simuladas situações sob H_0 completa para avaliar o erro tipo I do TB e o erro tipo I por experimento do 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. Foram consideradas populações com π_i 's idênticos e fixados em 0,01, 0,1 e 0,5, sob H_0 completa. Para a hipótese H_0 parcial, considerou-se uma diferença Δ entre os valores das proporções de dois grupos formados que internamente possuem proporções iguais. Essa diferença variou entre 0,01 e 0,9. Foram consideradas $k = 2, 5$ e 10 populações e tamanhos amostrais $n_i = 10, 30$ e 100 para cada população, $\forall i = 1, 2, \dots, k$. O TB proposto para a igualdade de várias proporções binomiais apresentou excelente desempenho e valores de poder relativamente altos. O TCMB para proporções binomiais, sob H_0 completa e parcial, foi conservativo e apresentou grandes valores de poder.

Palavras-chave: Simulação Monte Carlo, teste bayesiano de igualdade de várias proporções binomiais, teste comparações múltiplas bayesianas.

2 ABSTRACT

In frequentist theory, inferences about two or more 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 or to the need of implementing routines for their usage. 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 , $i = 1, 2, \dots, k$ were considered and 1.000 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 simulations were subdivided into two parts: the first one to test the null hypothesis of equality of the proportions denote TB and the second for performing the multiple comparisons bayesian tests (TCMB). Each one was again subdivided into two parts. First, there were simulated circumstances under complete H_0 to assess the type I error rate of TB and the experimentwise type I error rate of TCMB. In the second, circumstance under partial H_0 were simulated to evaluate the power of TB and TCMB, and also the experimentwise type I error rates for the TCMB. Populations with π_i 's identical and settled at 0.01, 0.1 and 0.5 were considered, under complete H_0 . For the partial null hypothesis H_0 , it was considered a difference Δ between the binomial proportions of two groups, where each one was considered internally homogeneous. This difference varied between 0.01 and 0.9. We considered $k = 2, 5$ and 10 populations and sample sizes $n_i = 10, 30$ and 100 for the i th population, $\forall i = 1, 2, \dots, k$. 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.

Key-words: Monte Carlo simulation, bayesian test for equality of several binomial proportions, bayesian multiple comparisons test.

3 Introdução

A comparação de várias proporções binomiais é relevante em muitos estudos científicos. Na inferência clássica, essa comparação é normalmente realizada 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 & 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. 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) e Krishnamoorthy & Peng (2008) avaliaram o desempenho de alguns desses testes, dos quais podem-se citar o teste de razão de verossimilhanças G^2 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. Em relação aos testes de comparações múltiplas, existem poucos trabalhos na literatura envolvendo proporções binomiais. Biase & Ferreira (2006) propuseram testes de comparações múltiplas para parâmetros binomiais utilizando métodos de reamostragem *bootstrap*, que apresentaram

excelentes resultados. O problema da aplicação dos testes de comparações múltiplas, como também dos testes assintóticos, está vinculado com a necessidade de implementação de rotinas para a realização desses testes que, para muitos pesquisadores, não é uma tarefa fácil.

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 homocedásticas e heterocedásticas (Andrade, 2008). No contexto de tabelas de contingência, Agresti & Min (2005) também avaliaram o desempenho de intervalos de credibilidade bayesiano para proporções binomiais via simulação Monte Carlo.

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 utilizando-se simulação Monte Carlo.

4 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 & Min (2005).

4.1 Teste bayesiano para igualdade de proporções binomiais

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

$$L(\mathbf{y}|\boldsymbol{\pi}) = \prod_{i=1}^k \frac{n_i!}{y_i!(n_i - y_i)!} \pi_i^{y_i} (1 - \pi_i)^{n_i - y_i}, \quad (4.1)$$

em que: $\mathbf{y} = [y_1, y_2, \dots, y_k]^\top$ é o vetor de dados e $\boldsymbol{\pi} = [\pi_1, \pi_2, \dots, \pi_k]^\top$ é o vetor de parâmetros das k populações binomiais independentes.

Optou-se por escolher *prioris* conjugadas. Logo, a distribuição 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:

$$p(\boldsymbol{\pi}) \propto \prod_{i=1}^k \pi_i^{\alpha_i - 1} (1 - \pi_i)^{\beta_i - 1}. \quad (4.2)$$

A distribuição conjunta *a posteriori* foi obtida multiplicando-se (4.1) por (4.2) e o resultado é:

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

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

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

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

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

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

$$E(\pi_i|\mathbf{y}) = \frac{\alpha_i + y_i}{\alpha_i + \beta_i + n_i} = \mu_{\pi_i}, \quad (4.3)$$

$$V(\pi_i|\mathbf{y}) = \frac{(\alpha_i + y_i)(\beta_i + n_i - y_i)}{(\alpha_i + \beta_i + n_i)^2(\alpha_i + \beta_i + n_i + 1)} = \sigma_{\pi_i}^2. \quad (4.4)$$

Sob a hipótese nula, $H_0 : \pi_1 = \pi_2 = \dots = \pi_k = \pi$, a função de verossimilhança

foi simplificada da seguinte forma:

$$\begin{aligned} L(\mathbf{y}|\pi) &\propto \pi^{y_1}(1-\pi)^{n_1-y_1}\pi^{y_2}(1-\pi)^{n_2-y_2}\dots\pi^{y_k}(1-\pi)^{n_k-y_k} \\ &\propto \pi^{\sum_{i=1}^k y_i}(1-\pi)^{\sum_{i=1}^k n_i - \sum_{i=1}^k y_i}. \end{aligned} \quad (4.5)$$

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

$$P(\pi) \propto \pi^{\alpha-1}(1-\pi)^{\beta-1}. \quad (4.6)$$

A distribuição *a posteriori* de π foi obtida multiplicando-se (4.5) por (4.6), da seguinte forma:

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

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

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

cuja média e variância são, respectivamente:

$$E(\pi|\mathbf{y}) = \frac{\alpha + \sum_{i=1}^k y_i}{\alpha + \beta + \sum_{i=1}^k n_i} = \mu_\pi, \quad (4.7)$$

$$V(\pi|\mathbf{y}) = \frac{\left(\alpha + \sum_{i=1}^k y_i\right) \left(\beta + \sum_{i=1}^k n_i - \sum_{i=1}^k y_i\right)}{\left(\alpha + \beta + \sum_{i=1}^k n_i\right)^2 \left(\alpha + \beta + \sum_{i=1}^k n_i + 1\right)} = \sigma_\pi^2. \quad (4.8)$$

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 (Ferreira, 2008; Johnson & Weerahandi, 1988). Assim, para realizar o teste será considerada a função dos parâmetros dada por:

$$q_c = (\boldsymbol{\pi}_p - \boldsymbol{\pi}_0)^\top \boldsymbol{\Sigma}_p^{-1} (\boldsymbol{\pi}_p - \boldsymbol{\pi}_0), \quad (4.9)$$

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

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

tendo $E(\pi_i|\mathbf{y})$ sido apresentada em (4.3), $\boldsymbol{\pi}_0$ é um vetor de médias de k cadeias independentes de π , sob H_0 , dado por:

$$\boldsymbol{\pi}_0 = \begin{bmatrix} \mu_\pi \\ \mu_\pi \\ \vdots \\ \mu_\pi \end{bmatrix}, \quad (4.10)$$

em que a média μ_π é dada em (4.7) 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:

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

em que $\sigma_{\pi_i}^2$ é dado em (4.4).

Como a distribuição da função paramétrica (4.9) é desconhecida, usou-se o procedimento descrito na sequência. Inicialmente, foram geradas k cadeias independentes da distribuição *a posteriori* de $\boldsymbol{\pi}$ 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 (Ferreira et al., 2005; Hinkelmann & Kempthorne, 1987; Steel & Torrie, 1980). Assim, foram gerados vetores $\boldsymbol{\pi}_j$, cujos componentes π_{ij} são realizações da distribuição *a posteriori* beta:

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

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 $j = 1, 2, \dots, N$ e $i = 1, 2, \dots, k$. 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 (4.10) e a matriz de covariâncias

Σ_0 é dada por:

$$\Sigma_0 = \begin{bmatrix} \sigma_\pi^2 & 0 & \dots & 0 \\ 0 & \sigma_\pi^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_\pi^2 \end{bmatrix} = \text{diag}(\sigma_\pi^2),$$

em que σ_π^2 é dado em (4.8).

Para cada unidade vetorial j da distribuição de equilíbrio *a posteriori* sob H_0 simulada (π_j), foi obtida a forma quadrática:

$$q_j = (\pi_j - \pi_0)^\top \Sigma_0^{-1} (\pi_j - \pi_0), \quad (4.11)$$

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

A quantidade (4.11) pode ser simplificada por:

$$q_j = \sum_{i=1}^k (\pi_{ij} - \mu_\pi)^2 \left[\frac{(\alpha + \beta + \sum_{i=1}^k n_i)^2 (\alpha + \beta + \sum_{i=1}^k n_i + 1)}{(\alpha + \sum_{i=1}^k y_i) (\beta + \sum_{i=1}^k n_i - \sum_{i=1}^k y_i)} \right]. \quad (4.12)$$

Os N valores de q_j , $j = 1, 2, \dots, N$ 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 (4.9) 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:

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

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.

4.2 Comparações múltiplas bayesianas

As comparações múltiplas são definidas para todos os testes de hipóteses entre duas proporções binomiais. Assim, para testar $H_0: \ell^\top \pi = 0$, em que o vetor ℓ 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 (4.12), propondo a seguinte função paramétrica:

$$\delta_c = \frac{\ell^\top \pi_p}{\sqrt{\ell^\top \Sigma_p^{-1} \ell}}.$$

Essa quantidade pode ser expandida na seguinte expressão:

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

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

$$\text{IC}_{1-\alpha}(\ell^\top \pi) : \ell^\top \pi_p \pm \sqrt{q_\alpha} \sqrt{\ell^\top \Sigma_p^{-1} \ell} \quad (4.13)$$

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

$$\text{RC}_{1-\alpha}(\boldsymbol{\pi}) : \{\boldsymbol{\pi} | (\boldsymbol{\pi} - \boldsymbol{\pi}_p)^\top \boldsymbol{\Sigma}_p^{-1} (\boldsymbol{\pi} - \boldsymbol{\pi}_p) \leq q_\alpha\}. \quad (4.14)$$

As expressões (4.13) e (4.14) não foram utilizadas diretamente na avaliação do teste de comparações múltiplas bayesiano (TCMB), mas possuem grande valor didático e científico.

4.3 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 $\pi_1, \pi_2, \dots, \pi_k$ e n_1, n_2, \dots, 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 (H_0 completa) 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 entre os valores dos parâmetros dos dois grupos considerados (Δ), 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 de Δ representa a diferença entre os dois grupos. 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 $\ell = 1, 2,$

3, 4, 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 $\Delta = 0,01, 0,10$ 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 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 ser não significativamente ($P < 0,01$) diferente do nível nominal, 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 $N = 1.000$ 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.

Considerando um nível nominal de probabilidade de 1% para o teste, observou-

se que, num total de $N = 1.000$ simulações para $\alpha = 5\%$, os valores inferiores a 33 e superiores a 69 levam à rejeição da hipótese nula $H_0 : \alpha = 5\%$ e que valores inferiores a 3 e superiores a 19 levam à rejeição da hipótese $H_0 : \alpha = 1\%$, considerando o mesmo nível de probabilidade de 1% para o teste realizado.

Os resultados do TCMB também foram comparados com os obtidos com os testes de *bootstrap*, apresentados por Biase & Ferreira (2006).

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 $i = 1, 2, \dots, k$ 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.

As simulações foram realizadas utilizando-se o *software* R (R. Development Core Team, 2008) e as rotinas implementadas para avaliação dos testes estão apresentadas em anexo.

5 Resultados e Discussão

5.1 Teste bayesiano

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 5.2 e 5.3.

5.2 Erro tipo I

5.2.1 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 2.18 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.

De modo geral, pode-se observar que houve controle do erro tipo I, pois não ocorreram situações em que os valores das 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 consi-

TABELA 2.18 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%.

k	n	$\pi = 0,01$	$\pi = 0,1$	$\pi = 0,5$
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 ^{ns}
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%.

^{ns} não significativamente diferente do nível nominal de 5%, considerando uma credibilidade de 99%.

derando 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 2.19 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%.

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

TABELA 2.19 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%.

k	n	$\pi = 0,01$	$\pi = 0,1$	$\pi = 0,5$
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 ^{ns}
5	30	0,00 ⁺	3,10 ⁺	3,50 ^{ns}
5	100	0,00 ⁺	4,20 ^{ns}	3,60 ^{ns}
10	10	0,00 ⁺	0,10 ⁺	8,20 [*]
10	30	0,00 ⁺	2,50 ⁺	5,00 ^{ns}
10	100	0,00 ⁺	4,50 ^{ns}	3,30 ^{ns}

* significativamente superior ao nível nominal de 5%, considerando uma credibilidade de 99%.

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

^{ns} não significativamente diferente do nível nominal de 5%, considerando uma credibilidade de 99%.

obtidos das taxas de erro tipo I, considerando os hiperparâmetros iguais a 1 e 2. Agresti & 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.

Comparando-se os resultados das taxas de erro tipo I sob H_0 completa do TB (Tabela 2.18) com os obtidos para os testes G^2 e X^2 (Tabela 1.2), verifica-se que o TB é muito mais conservativo, principalmente em relação ao teste G^2 . Provavelmente, os valores de poder do TB serão inferiores aos valores de poder dos testes G^2 e X^2 .

5.3 Poder

5.3.1 Poder sob H_0 parcial

Houve similaridade muito grande entre o desempenho do TB simulado, considerando o nível nominal de 1% e 5%. Por essa razão, serão apresentados apenas os resultados para 5%.

Na Tabela 2.20, são apresentados, em porcentagem, os valores de 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, independente do número de populações (k) e do tamanho amostral (n).

Para diferenças pequenas ou moderadas ($0,1 \leq \Delta \leq 0,3$), observa-se que os valores de poder do TB são pequenos para todos os casos, considerando pequenas amostras ($n = 10$) e, em alguns casos, 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 os valores de 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.

TABELA 2.20 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
5	10	0,00	0,00	0,00	0,90	6,90	29,90
5	30	0,00	0,00	0,90	42,60	92,30	99,90
5	100	0,00	9,90	87,60	100,00	100,00	100,00
10	10	0,00	0,00	0,00	2,00	19,30	64,30
10	30	0,00	0,00	1,10	86,20	99,90	100,00
10	100	0,00	30,00	99,80	100,00	100,00	100,00

		Δ				
k	n	0,5	0,6	0,7	0,8	0,9
5	10	62,00	88,70	98,60	100,00	100,00
5	30	100,00	100,00	100,00	100,00	100,00
5	100	100,00	100,00	100,00	100,00	100,00
10	10	95,70	100,00	100,00	100,00	100,00
10	30	100,00	100,00	100,00	100,00	100,00
10	100	100,00	100,00	100,00	100,00	100,00

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 2.21 são apresentados esses valores de 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 os valores de poder do TB considerando todos os hiperparâmetros iguais a 2. No entanto, observa-se que, na maioria das situações, os valores de poder do TB com hiperparâmetros iguais a 1 e 2 foram superiores aos valores de poder do TB, considerando apenas hiperparâmetros iguais a 2.

Comparando-se esses resultados, pode-se observar que, para $\Delta = 0,01$, os valores de poder do TB são relativamente iguais para as duas configurações de hiperparâmetros considerados, independente do número de populações (k) e do tamanho amostral (n). Para valores de Δ maiores ($\Delta = 0,05$), os valores de poder do TB, considerando

TABELA 2.21 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
5	10	0,00	0,00	0,10	2,50	20,60	57,00
5	30	0,00	0,10	7,90	77,00	99,50	100,00
5	100	0,00	45,00	97,30	100,00	100,00	100,00
10	10	0,00	0,00	0,00	4,20	41,90	85,60
10	30	0,00	0,00	15,70	98,00	100,00	100,00
10	100	0,10	77,00	100,00	100,00	100,00	100,00

		Δ				
k	n	0,5	0,6	0,7	0,8	0,9
5	10	82,50	96,50	99,80	100,00	100,00
5	30	100,00	100,00	100,00	100,00	100,00
5	100	100,00	100,00	100,00	100,00	100,00
10	10	99,20	99,80	100,00	100,00	100,00
10	30	100,00	100,00	100,00	100,00	100,00
10	100	100,00	100,00	100,00	100,00	100,00

os hiperparâmetros iguais a 1 e 2, são expressivamente superiores aos de 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 os valores de poder foram consideravelmente superiores ao do TB, com hiperparâmetros iguais a 2. Para Δ grande ou muito grande ($\Delta \geq 0,4$), houve casos de superioridade dos valores de 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.

Outra comparação que deve ser realizada refere-se aos valores de poder do TB

com os dos testes G^2 e X^2 . Analisando-se as Tabelas 2.20 e 1.4, o que se verifica é que os valores de poder do TB com hiperparâmetros iguais a 2 são inferiores aos valores de poder dos testes G^2 e X^2 , principalmente se $\Delta \leq 0,3$. À medida que os valores de Δ aumentam, a diferença entre os valores de poder destes testes diminui, pois o poder de todos os testes tende a se aproximar de 100%. Esse mesmo desempenho foi observado ao comparar-se o TB, considerando hiperparâmetros α_i 's e β_i 's iguais a 1 e α 's e β 's iguais a 2 (Tabela 2.21), em relação aos testes G^2 e X^2 (Tabela 1.4). A inferioridade do poder do TB já era esperada, pois, conforme comentado anteriormente, o TB apresentou taxas de erro tipo I mais conservativas em relação aos demais testes avaliados neste trabalho.

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, considerando o nível nominal de 1% e 5%. Novamente, os resultados para $\alpha = 1\%$ foram bastante similares aos observados para $\alpha = 5\%$ e, por isso, foram apresentados apenas os resultados para $\alpha = 5\%$.

Na Tabela 2.22 são apresentados os valores de 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 os valores de poder do TB são inferiores ao nível nominal de 5% para todos os tamanhos amostrais (n), número de populações (k) e valores de $\pi^{(1)}$.

Considerando $\Delta = 0,1$, os valores de poder ainda são pequenos, no entanto, nota-se que o poder do TB tem um crescimento à medida que o número de populações e tamanhos amostrais 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, os valores de poder tenderam a se aproximar de 100%.

TABELA 2.22 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%.

k	$\pi^{(1)}$	n	$\Delta = 0,01$	$\Delta = 0,1$	$\Delta = 0,4$
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

Observa-se que os valores de poder do TB com hiperparâmetros iguais a 2 também foram inferiores aos valores de poder dos testes G^2 e X^2 (Tabelas 2.22 e 1.6), independente dos valores de k , n , $\pi^{(1)}$, para $\Delta = 0,01$ e $0,1$. Para $\Delta = 0,4$, o desempenho do TB e o dos testes G^2 e X^2 tendem a se igualar e a se aproximar de 100%, principalmente se $n \geq 30$.

5.4 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.

5.5 Erro tipo I

5.5.1 Erro tipo I sob H_0 completa

Na Tabela 2.23 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.

Comparando-se esses resultados com os obtidos para o TCM utilizando a distribuição assintótica de qui-quadrado (Tabela 1.8) e também com os testes de *bootstrap* avaliados por Biase (2006) (Tabela 1.1), observa-se que o TCMB é extremamente conservativo em relação aos demais testes, 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 2.24, 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

TABELA 2.23 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%.

k	n	$\pi = 0,01$	$\pi = 0,1$	$\pi = 0,5$
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 ^{ns}
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%.

^{ns} não significativamente diferente do nível nominal de 5%, considerando uma credibilidade de 99%.

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.

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.

Os resultados das taxas de erro tipo I, sob H_0 completa, do TCMB considerando $\alpha = 1\%$, foram bastante similares aos resultados observados para $\alpha = 5\%$ e, em decorrência desse fato, não foram apresentados.

TABELA 2.24 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%.

k	n	$\pi = 0,01$	$\pi = 0,1$	$\pi = 0,5$
2	10	0,00 ⁺	0,50 ⁺	4,60 ^{ns}
2	30	0,00 ⁺	1,80 ⁺	3,90 ^{ns}
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%.

^{ns} não significativamente diferente do nível nominal de 5%, considerando uma credibilidade de 99%.

5.5.2 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 2.25, 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.

Comparando-se o desempenho do TCMB com o TCM baseado na distribuição assintótica, quanto às taxas de erro tipo I sob H_0 parcial, pode-se observar que o desempenho do TCMB assemelha-se ao do TCM assintótico, pois ambos os testes foram conservativos em todas as situações avaliadas. No entanto, as taxas de erro tipo I do TCMB foram expressivamente menores do que as do TCM, o que pode afetar o poder.

Os testes de *bootstrap* avaliados por Biase (2006) apresentaram melhor de-

TABELA 2.25 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%.

k	n	Δ					
		0,01	0,05	0,1	0,2	0,3	0,4
5	10	0,00 ⁺					
5	30	0,00 ⁺					
5	100	0,00 ⁺	0,00 ⁺	0,00 ⁺	0,10 ⁺	0,10 ⁺	0,00 ⁺
10	10	0,00 ⁺					
10	30	0,00 ⁺	0,00 ⁺	0,00 ⁺	0,00 ⁺	0,10 ⁺	0,10 ⁺
10	100	0,00 ⁺	0,10 ⁺				

k	n	Δ				
		0,5	0,6	0,7	0,8	0,9
5	10	0,00 ⁺				
5	30	0,00 ⁺	0,10 ⁺	0,20 ⁺	0,00 ⁺	0,00 ⁺
5	100	0,20 ⁺	0,00 ⁺	0,20 ⁺	0,10 ⁺	0,00 ⁺
10	10	0,00 ⁺				
10	30	0,00 ⁺				
10	100	0,00 ⁺				

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

sempenho, pois, em alguns casos, o tamanho dos testes não foi significativamente ($P < 0,01$) diferente do valor nominal adotado. Porém, conforme mencionado anteriormente, a aplicação destes testes ainda é limitada por necessitar de conhecimento de programação e de grande esforço computacional.

Na Tabela 2.26 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 2.26 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
5	10	0,00 ⁺	0,00 ⁺	0,00 ⁺	0,10 ⁺	0,90 ⁺	0,80 ⁺
5	30	0,00 ⁺	0,00 ⁺	0,00 ⁺	0,00 ⁺	0,20 ⁺	0,20 ⁺
5	100	0,00 ⁺	0,00 ⁺	0,00 ⁺	0,20 ⁺	0,30 ⁺	0,10 ⁺
10	10	0,00 ⁺	0,10 ⁺	0,00 ⁺	0,10 ⁺	1,00 ⁺	2,10 ⁺
10	30	0,00 ⁺	0,00 ⁺	0,00 ⁺	0,10 ⁺	0,10 ⁺	0,20 ⁺
10	100	0,00 ⁺	0,00 ⁺	0,20 ⁺	0,10 ⁺	0,00 ⁺	0,00 ⁺

		Δ				
k	n	0,5	0,6	0,7	0,8	0,9
5	10	1,10 ⁺	1,40 ⁺	0,60 ⁺	0,10 ⁺	0,00 ⁺
5	30	0,40 ⁺	0,20 ⁺	0,50 ⁺	0,00 ⁺	0,00 ⁺
5	100	0,10 ⁺	0,20 ⁺	0,30 ⁺	0,00 ⁺	0,10 ⁺
10	10	2,30 ⁺	1,50 ⁺	0,80 ⁺	0,40 ⁺	0,00 ⁺
10	30	0,10 ⁺	0,10 ⁺	0,10 ⁺	0,00 ⁺	0,00 ⁺
10	100	0,00 ⁺	0,00 ⁺	0,10 ⁺	0,00 ⁺	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.

Para o valor nominal de 1%, os resultados das taxas de erro tipo I por experimento foram semelhantes aos observados para 5%. Assim, a maioria dos resultados foi significativamente ($P < 0,01$) inferior a $\alpha = 1\%$, existindo apenas alguns casos em que os resultados do TCMB não diferiram significativamente ($P > 0,01$) de 1%, considerando hiperparâmetros iguais a 0,01 e 2.

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 2.27, em função de k , $\pi^{(1)}$ e n .

TABELA 2.27 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%.

k	$\pi^{(1)}$	n	$\Delta = 0,01$	$\Delta = 0,1$	$\Delta = 0,4$
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 2.25) 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.

5.6 Poder

5.6.1 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. Na Tabela 2.28 são apresentados esses valores de poder do TCMB com hiperparâmetros iguais a 0,01 e 2 em função de k , n e Δ , considerando $\alpha = 5\%$.

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, $0,1 \leq \Delta \leq 0,3$, os valores de poder tendem 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 ($0,4 \leq \Delta \leq 0,6$), 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 ($\Delta \geq 0,7$), os valores de poder são pequenos 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 dos valores de poder do teste, ao contrário do que ocorre com o aumento do número de populações (k), que propicia a redução dos valo-

TABELA 2.28 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
5	10	0,00	0,00	0,00	0,85	6,40	17,95
5	30	0,00	0,03	0,63	16,46	59,60	88,80
5	100	0,00	1,76	35,30	97,20	100,00	100,00
10	10	0,00	0,02	0,00	0,12	1,35	6,18
10	30	0,00	0,00	0,00	1,02	15,01	53,99
10	100	0,00	0,00	3,66	76,10	99,42	100,00

		Δ				
k	n	0,5	0,6	0,7	0,8	0,9
5	10	39,78	67,76	87,03	97,00	99,83
5	30	98,50	100,00	100,00	100,00	100,00
5	100	100,00	100,00	100,00	100,00	100,00
10	10	18,04	38,57	64,83	86,68	98,78
10	30	87,05	98,24	99,91	100,00	100,00
10	100	100,00	100,00	100,00	100,00	100,00

res de poder. Essas informações estão de acordo com as obtidas para o poder do TCM utilizando distribuição assintótica de qui-quadrado que estão apresentadas na Tabela 1.13. Entretanto, observa-se que, para $\Delta \leq 0,4$, o TCMB apresentou valores de poder inferiores ao TCM na maioria das situações avaliadas. Mas, à medida que a diferença Δ tornou-se maior ($\Delta \geq 0,5$), a maioria dos valores de poder do TCMB aproximou-se de 100% e igualou-se aos valores de poder dos testes G^2 e X^2 .

Para o nível nominal de 1%, observou-se que o TCMB com hiperparâmetros iguais a 0,01 e 2 apresentou desempenho similar aos observados para $\alpha = 5\%$. Os valores de poder do TCMB em função de k , n e Δ , considerando $\alpha = 1\%$, são apresentados na Tabela 2.29. Verifica-se, para $\Delta \leq 0,1$, que os valores de poder são relativamente próximos de zero.

TABELA 2.29 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 1%.

		Δ					
k	n	0,01	0,05	0,1	0,2	0,3	0,4
5	10	0,00	0,00	0,00	0,70	5,27	16,26
5	30	0,00	0,00	0,00	2,65	27,03	69,70
5	100	0,00	0,05	0,95	87,45	99,83	100,00
10	10	0,00	0,02	0,00	0,12	1,34	6,06
10	30	0,00	0,00	0,00	0,10	6,98	37,53
10	100	0,00	0,00	0,36	48,34	97,14	99,99

		Δ				
k	n	0,5	0,6	0,7	0,8	0,9
5	10	36,30	62,83	82,67	94,83	99,48
5	30	94,08	99,50	99,98	100,00	100,00
5	100	100,00	100,00	100,00	100,00	100,00
10	10	17,54	37,41	62,63	83,99	97,07
10	30	76,98	95,64	99,72	100,00	100,00
10	100	100,00	100,00	100,00	100,00	100,00

Para uma diferença moderada de Δ ($0,2 \leq \Delta \leq 0,4$), os valores de poder do TCMB foram pequenos para amostras pequenas e intermediárias ($n \leq 30$). Com o aumento de n de 30 para 100, houve um crescimento expressivo dos valores de poder, aproximando-se de 100%, tanto para $k = 5$ como para $k = 10$. Isso não foi verificado apenas para $k = 10$, $n = 100$ e $\Delta = 0,2$.

Considerando valores de Δ grandes ou muito grandes ($\Delta \geq 0,5$), os valores de poder são pequenos (inferiores a 70%) para $k = 5$ e 10 e $n = 10$, considerando $\Delta = 0,5$ e 0,6 e com $k = 10$ e $n = 10$ para $\Delta = 0,7$. Nos demais casos, os valores de poder aproximam-se de 100%, mesmo para pequenas amostras ($n = 10$).

De maneira geral, pode-se observar que houve redução do poder do TCMB com o aumento de k de 5 para 10 e que, com o aumento do tamanho das amostras, o poder

dos testes tende a aumentar. Esses resultados estão de acordo com os apresentados para o TCM (Tabela 1.14) e também com os testes de *bootstrap* avaliados por Biase (2006).

Comparando-se os resultados das Tabelas 2.28 e 2.29 com os das Tabelas 1.13 e 1.14, respectivamente, é possível verificar que, tanto para $\alpha = 1\%$ como para $\alpha = 5\%$, o TCM apresentou, na grande maioria das situações, valores de poder inferiores aos valores de poder do TCMB, considerando diferenças pequenas e moderadas de Δ ($\Delta \leq 0,4$). Para diferenças grandes ou muito grandes, $\Delta \geq 0,5$, o TCMB apresentou desempenho melhor, tendo valores de poder superiores aos do TCM.

Quanto aos testes de *bootstrap*, verifica-se, para $\Delta = 0,1$, que os valores de poder do TCMB são menores para $n \leq 30$ e superiores aos testes de *bootstrap*, para $n = 100$. Considerando $\Delta = 0,5$, os testes de *bootstrap* apresentaram valores de poder superiores ao TCMB, independente do tamanho das amostras. Por fim, para $\Delta = 0,9$, o teste *bootstrap* Pan apresentou desempenho semelhante ao do TCMB, enquanto o teste *bootstrap* MV apresentou valores de poder expressivamente inferiores ao TCMB para $n \leq 30$. Para grandes amostras ($n = 100$), o poder dos dois testes aproximou-se de 100%, igualando-se.

6 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.

7 Referências Bibliográficas

- AGRESTI, A.; COULL, B.A. Approximate is better than “exact” for interval estimation of binomial proportions. **American Statistician**, Alexandria, v.52, n.2, p.119-126, May 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, June 2005.
- ANDRADE, P.C.R. **Comparações múltiplas bayesianas em modelos normais homocedásticos e heterocedásticos**. 2008. 96p. Tese (Doutorado em Estatística) - Universidade Federal de Lavras, Lavras.
- BIASE, N.G. **Comparações múltiplas para parâmetros binomiais utilizando bootstrap**. 2006. 73p. Dissertação (Mestrado em Agronomia) - Universidade Federal de Lavras, Lavras.
- BIASE, N.G.; FERREIRA, D.F. Comparações múltiplas para proporções binomiais utilizando bootstrap. **Revista de Matemática e Estatística**, São Paulo, v.24, n.1, p.95-112, 2006.
- FERREIRA, D.F.; DEMÉTRIO, C.G.B.; MANLY, B.F.J.; MACHADO, A.A. Aplicações dos métodos de bootstrap nos procedimentos de comparações múltiplas. In: REUNIÃO ANUAL DA REGIÃO BRASILEIRA DA SOCIEDADE INTERNACIONAL DE BIOMETRIA, 50., 2005, Londrina. **Anais...** Londrina: SEAGRO, 2005, 5p.
- FERREIRA, D.F. **Estatística multivariada**. Lavras: UFLA, 2008. 662p. 1 v.
- HINKELMANN, K.; KEMPTHORNE, O. **Design and analysis of experiments**. New York: J. Wiley, 1987. 445p. 1 v.
- JOHNSON, R.A.; WEERAHANDI, S. A bayesian solution to the multivariate Behrens-Fisher problem. **Journal of the American Statistical Association**, 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. **Journal of Applied Statistical Science**, 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. **Computational Statistics & Data Analysis**, Amsterdam, v.45, p.697-707, 2004.

NELDER, J.A.; WEDDERBURN, R.W.M. Generalized linear models. **Journal of the Royal Statistical Society**. Series A. Statistics in society, London, v.135, p.370-384, 1972.

R Development Core Team. **R**: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, 2008. Disponível em: <<http://www.r-project.org>>. Acesso em: 10 out. 2008.

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. **Journal of the Royal Statistical Society**, London, v.37, n.3, p.421-434, 1988.

ANEXOS

	Página
PROGRAMA A: Programa R de simulação utilizado para computar o erro tipo I, erro tipo I por experimento e o poder de todos os testes, sob H_0 completa e parcial, conforme a situação avaliada	138
PROGRAMA B: Programa R para realizar inferências sobre proporções binomiais	148

PROGRAMA A: Programa R de simulação utilizado para computar o erro tipo I, erro tipo I por experimento e o poder de todos os testes, sob H_0 completa e parcial, conforme a situação avaliada.

```
# Programa para simular populações binomiais independentes
# (ni, pi) i=1,2,...,k e para avaliar o desempenho de testes da
# hipótese H0: p1 = p2 = ... = pk e de comparações múltiplas
# (poder e taxa de erro tipo I por experimento), utilizando
# inferências frequentistas e bayesianas.

# Função para simular k populações binomiais independentes de
# tamanho nis e probabilidades de sucesso pis – retorna o vetor
# y=[y1, ..., yk].

gerabin <- function(k,nis,pis) {
  if ((length(nis)!=length(pis)) | (length(nis)!=k))
    stop("Dimensões incompatíveis!")
  return(rbinom(k,nis,pis))
}

# Função que retorna a média e a matriz de covariâncias da
# posteriori sob H1, referente as expressões (4.4) e (4.5) da
# tese, recebendo: o vetor y (k x 1) de observações, o vetor nis
# (k x 1) dos tamanhos amostrais, o número k de populações, o
# vetor alphas de hiperparâmetros alpha_is e o vetor betas dos
# hiperparâmetros beta_is (k x 1).

MedCovPostH1 <- function(k,nis,y,alphas,betas)
{
  pip <- (alphas+y)/(alphas+betas+nis)
  sigp <- (alphas+y)*(betas+nis-y)/((alphas+betas+nis)^2*
    (alphas+betas+nis+1))
  return(list(pip=pip,sigp=diag(sigp)))
}

# Função que retorna a média e a matriz de covariâncias da
# posteriori sob H0, referente as expressões (4.9) e (4.10) da
# tese, recebendo: o vetor y (k x 1) de observações, o vetor nis
# (k x 1) dos tamanhos amostrais, o número k de populações, o
# hiperparâmetro alpha e o hiperparâmetro beta.

MedCovPostH0 <- function(k,nis,y,alpha,beta) {
  pi0 <- rep((alpha+sum(y))/(alpha+beta+sum(nis)),times=k)
  sig0 <- (alpha+sum(y))*(beta+sum(nis)-sum(y))/
    ((alpha+beta+sum(nis))^2*(alpha+beta+sum(nis)+1))
  sig0 <- rep(sig0,times=k)
  return(list(pi0=pi0,sig0=diag(sig0)))
}
```

```

# Função para obter a distribuição das formas quadráticas q_j sob
# H0, representada pela equação (4.18) da tese. Deve receber os
# hiperparâmetros alpha e beta, os tamanhos amostrais nis, as
# observações y (k x 1) e o número de populações k. Retorna o vetor
# de quantis (qalpha) (q x 1), dado alph[alpha1, alpha2, ..., alphaq]
# e N simulações.

quantisdeqj <- function(k, nis, y, alpha, beta, alph, N=50000) {
  pi0cov0 <- MedCovPostH0(k, nis, y, alpha, beta)
  alphap0 <- alpha + sum(y)
  betap0 <- beta + sum(nis) - sum(y)
  pi0p <- (matrix(rbeta(k*N, alphap0, betap0), k, N) -
    pi0cov0$pi0)^2
  pi0p <- pi0p/pi0cov0$sig0[1,1] # artifício
  pi0p <- sort(apply(pi0p, 2, sum))
  qalpha <- round((1 - alph)*N)
  qalpha[qalpha==0] <- 1
  qalpha <- pi0p[qalpha]
  return(qalpha)
}

# Função para aplicar o teste bayesiano (TB) para a hipótese
# H0: p1 = p2 = ... = pk com hiperparâmetros alphas, betas, alpha
# e beta. Deve receber os quantis qalpha sob H0 da distribuição
# de qjs.

BTBin <- function(k, nis, y, alphas, betas, alpha, beta, qalpha) {
  pi0cov0 <- MedCovPostH0(k, nis, y, alpha, beta)
  pilcov1 <- MedCovPostH1(k, nis, y, alphas, betas)
  sig0inv <- diag(1/diag(pilcov1$sigp))
  qc <- t(pilcov1$pi0 - pi0cov0$pi0)%%sig0inv%%
  (pilcov1$pi0 - pi0cov0$pi0)
  sig <- rep(qc, times=length(qalpha)) > qalpha
  # TRUE para significativo
  return(list(qc=qc, sig=sig))
}

# Função para aplicar o teste de razão de verossimilhanças (G2)
# da equação (4.11) da tese para testar a hipótese H0: p1 = p2 =
# ... = pk.

LRTBin <- function(k, nis, y) {
  hatpis <- y/nis
  n <- sum(nis)
  somaYj <- sum(y)
  hatpi0 <- sum(y)/n
  aux <- any(y>0)
  if (aux==FALSE) aux1 <- n*y else
  aux1 <- n*y/(nis*somaYj)
}

```

```

aux      <- aux1>0
aux1     <- sum(y[aux]*log(aux1[aux]))
aux2     <- (nis-y)
aux      <- aux2>0
aux2     <- sum(aux2[aux]*log(n*aux2[aux]/(nis[aux]*(n-somaYj))))
chi1     <- 2*(aux1+aux2)
df       <- k - 1
pr.chi   <- pchisq(chi1,df,lower.tail=FALSE)
return(list(hatpis=hatpis,hatpi0=hatpi0,chi=chi1,
df=df,pr.chi=pr.chi))
}

# Função para aplicar o teste de formas quadráticas(X2) da equação
# (4.13) da tese para testar a hipótese H0: p1 = p2 = ... = pk.

QFTBin   <- function(k,nis,y) {
  hatpi0  <- sum(y)/sum(nis)
  aux1    <- (y-nis*hatpi0)^2
  aux2    <- nis*hatpi0*(1-hatpi0)
  aux     <- !any(aux2<=0)
  if (aux==TRUE) chi2 <- sum(aux1/aux2) else chi2 <- 0
  df      <- k - 1
  pr.chi  <- pchisq(chi2,df,lower.tail=FALSE)
  return(list(hatpi0=hatpi0,chi=chi2,df=df,pr.chi=pr.chi))
}

# Função para aplicar o teste de comparações múltiplas (TCM),
# representado pela equação (5.2) da tese. Retorna uma matriz com
# valores na diagonal dos estimadores de pis. Acima da diagonal
# tem-se o valor da estatística (pi vs pj), i diferente de j e
# variando de 1 até k, e abaixo da diagonal os valores-p referente
# à estatística.

MCTBin   <- function(k,nis,y) {
  res     <- diag(y/nis)
  hatpi0  <- sum(y)/sum(nis)
  varpi   <- hatpi0*(1-hatpi0)/nis
  df      <- k - 1
  for (i in 1:(k-1))
  {
    for (j in (i+1):k)
    {
      if ((varpi[i]==0) & (varpi[j]==0))
      {
        res[i,j] <- 0
      } else
      res[i,j] <- (res[i,i]-res[j,j])^2/(varpi[i]+varpi[j])
      res[j,i] <- pchisq(res[i,j],df,lower.tail=FALSE)
    }
  }
}

```

```

    return(res)
  }

# Função que recebe o valor de um objeto de MCTBin (res) e retorna
# se houve pelo menos uma significância para um vetor de níveis
# alpha

MCTETIEBin <- function(res , alpha=c(0.05 ,0.01)) {
  ct <- c(0,0)
  if (any(res[lower.tri(res)]<=alpha[1])) ct[1] <- 1
  if (any(res[lower.tri(res)]<=alpha[2])) ct[2] <- 1
  return(ct)
}

# Função para aplicar o teste de comparações múltiplas TCM),
# representado pela equação (5.2) da tese. Retorna uma matriz com
# valores na diagonal dos estimadores de pis , acima da diagonal o
# valor da estatística (pi vs pj) e abaixo da diagonal os valores-
# p referente à estatística. Retorna ctH0 indicando se houve ou
# não pelo menos 1 erro do tipo I e ctH1 indicando o número de
# significâncias entre grupos - somente para simulação sob H0
# parcial.

MCTBinPar <- function(k,nis ,y, alpha=c(0.05 ,0.01)) {
  res <- diag(y/nis)
  hatpi0 <- sum(y)/sum(nis)
  varpi <- hatpi0*(1-hatpi0)/nis
  df <- k - 1
  ctETI <- c(0,0)
  ctPod <- c(0,0)
  k1 <- k %/% 2
  if (k %/% 2 != 0) k1 <- k1 + 1
  k2 <- k - k1
  achei5 <- FALSE
  achei1 <- FALSE
  for (i in 1:(k-1))
  {
    for (j in (i+1):k)
    {
      if ((varpi[i]==0) & (varpi[j]==0))
      {
        res[i,j] <- 0
      } else
      res[i,j] <- (res[i,i]-res[j,j])^2/(varpi[i]+varpi[j])
      res[j,i] <- pchisq(res[i,j],df,lower.tail=FALSE)
      if (((i<=k1) & (j<=k1)) | ((i>k1) & (j>k1)))
      {
        if (res[j,i]<=alpha[1]) achei5=TRUE
        if (res[j,i]<=alpha[2]) achei1=TRUE
      } else
    }
  }
}

```

```

    {
      if (res[j,i]<=alpha[1]) ctPod[1] <- ctPod[1] + 1
      if (res[j,i]<=alpha[2]) ctPod[2] <- ctPod[2] + 1
    }
  }
}
if (achei5==T) ctETI[1] <- 1
if (achei1==T) ctETI[2] <- 1
return(list(res=res, ctETI=ctETI, ctPod=ctPod))
}

# Função para aplicar o teste de comparações múltiplas
# bayesiano(TCMB) da tese. Retorna um vetor com valores na diagonal
# das médias da posteriori de pis sob H1 e outro vetor com o valor
# da estatística qc (pi vs pj) qcalpha1 é uma matriz com TRUE=1 ou
# FALSE=0 (se significativo ou não) abaixo e o valor da estatística
# acima da diagonal, idem qcalpha2 para o outro alpha-estatística.

MCTBinBayes <- function(k, nis, y, alphas, betas, alpha, beta,
  alph=c(0.05,0.01)) {
  pilcov1 <- MedCovPostH1(k, nis, y, alphas, betas)
  qalpha <- quantisdeqj(k, nis, y, alpha, beta, alph, 10000)
  qcalpha1 <- matrix(0,k,k)
  qcalpha2 <- matrix(0,k,k)
  for (i in 1:(k-1))
  {
    elle <- matrix(0,k,1)
    elle[i] <- 1
    for (j in (i+1):k)
    {
      elle[j] <- -1
      if ((pilcov1$sigp[i,i]==0) & (pilcov1$sigp[j,j]==0))
      {
        qcalpha1[i,j] <- 0
        qcalpha2[i,j] <- 0
      } else
      {
        aux <- pilcov1$sigp[i,i]+pilcov1$sigp[j,j]
        qcalpha1[i,j] <- (t(elle)%*%pilcov1$pip)/sqrt(aux)
        qcalpha2[i,j] <- qcalpha1[i,j]
        if (abs(qcalpha1[i,j]) >= sqrt(qalpha[1]))
          qcalpha1[j,i] <- TRUE else qcalpha1[j,i] <- FALSE
        if (abs(qcalpha2[i,j]) >= sqrt(qalpha[2]))
          qcalpha2[j,i] <- TRUE else qcalpha2[j,i] <- FALSE
      }
    }
    elle[j] <- 0
  }
}
return(list(qc1=qcalpha1, qc2=qcalpha2))
}

```

```

# Função que recebe o valor de um objeto de MCTBinBayes e retorna
# se houve pelo menos uma significância para um vetor de níveis
# alpha qc$qc1 para primeiro alpha e qc$qc2 para o segundo alpha.

MCTETIEBinBayes <- function(qc) {
  ct <- c(0,0)
  if (any(qc$qc1[lower.tri(qc$qc1)]==1)) ct[1] <- 1
  if (any(qc$qc2[lower.tri(qc$qc2)]==1)) ct[2] <- 1
  return(ct)
}

# Função para aplicar o teste de comparações múltiplas bayesiano
# (TCMB), da tese sob H0 parcial. Retorna ctho indicando se houve
# ou não pelo menos 1 erro do tipo I e cth1 indicando o número de
# significâncias entre grupos – somente para simulação H0 parcial.

MCTBinParBayes <- function(k, nis, y, alphas, betas, alpha, beta,
  alph=c(0.05,0.01)){
  qc <- MCTBinBayes(k, nis, y, alphas, betas, alpha, beta, alph)
  ctETI <- c(0,0)
  ctPod <- c(0,0)
  k1 <- k %/% 2
  if (k %/% 2 != 0) k1 <- k1 + 1
  k2 <- k - k1
  achei5 <- FALSE
  achei1 <- FALSE
  for (i in 1:(k-1))
  {
    for (j in (i+1):k)
    {
      if (((i<=k1) & (j<=k1)) | ((i>k1) & (j>k1)))
      {
        if (qc$qc1[j,i]==1) achei5=TRUE
        if (qc$qc2[j,i]==1) achei1=TRUE
      } else
      {
        if (qc$qc1[j,i]==1) ctPod[1] <- ctPod[1] + 1
        if (qc$qc2[j,i]==1) ctPod[2] <- ctPod[2] + 1
      }
    }
  }
  if (achei5==T) ctETI[1] <- 1
  if (achei1==T) ctETI[2] <- 1
  return(list(ctETI=ctETI, ctPod=ctPod))
}

# Avaliação dos testes (G2 e X2) para o erro tipo I sob H0
# completa e o poder sob H0 parcial.

```

```

# Delta é a diferença entre proporções binomiais dos grupos ,
# especificado apenas sob H0 parcial.

# nkisgl é o número de proporções pertencentes ao primeiro grupo ,
# conforme número de populações.

# pi1 é o valor da proporção do primeiro grupo nis são os
# tamanhos amostrais.

k      <- 2
Delta  <- 0.0
nkisgl <- 2
pi1    <- 0.01
nis    <- rep(c(10),k)
pis    <- c(rep(c(pi1), nkisgl), rep(c(pi1+Delta), k-nkisgl))
nsim   <- 10000
nsr = matrix(0,2,2)
rownames(nsr) = c("LRT", "Formas Quadráticas")
colnames(nsr)=c("alpha=0.05", "alpha=0.01")
for (i in 1:nsim)
{
  y      <- gerabin(k, nis , pis)
  chi1 <- LRTBin(k, nis , y)
  chi2 <- QFTBin(k, nis , y)
  if (chi1$pr.chi <=0.05) nsr[1,1] <- nsr[1,1] + 1/nsim
  if (chi1$pr.chi <=0.01) nsr[1,2] <- nsr[1,2] + 1/nsim
  if (chi2$pr.chi <=0.05) nsr[2,1] <- nsr[2,1] + 1/nsim
  if (chi2$pr.chi <=0.01) nsr[2,2] <- nsr[2,2] + 1/nsim
}
nsr

# Avaliação do teste de comparações múltiplas(TCM) especificamente
# para o erro tipo I por experimento sob H0 completa

k      <- 2
nis    <- rep(c(10),k)
pis    <- rep(c(0.01),k)
nsim   <- 10000
etIe = matrix(0,1,2)
rownames(etIe) = c("Erro Tipo I Por Experimento")
colnames(etIe)=c("alpha=0.05", "alpha=0.01")
for (i in 1:nsim)
{
  y      <- gerabin(k, nis , pis)
  res    <- MCTBin(k, nis , y)
  ct     <- MCTETIEBin(res)
  etIe[1,] <- etIe[1,] + ct/nsim
}
etIe

```

```

# Avaliação do teste de comparações múltiplas(TCM) especificamente
# para o erro tipo I por experimento e poder sob H0 parcial.

k      <- 5
k1     <- k %% 2
if (k %% 2 != 0) k1 <- k1 + 1
k2     <- k - k1
nis    <- rep(c(10),k)
pis    <- c(rep(c(0.01),k1),rep(c(0.02),k2))
nsim   <- 10000
etIePod = matrix(0,2,2)
rownames(etIePod) = c("ETI Por Experimento - H0 parcial",
"Poder - H0 parcial")
colnames(etIePod)=c("alpha=0.05","alpha=0.01")

for (i in 1:nsim)
{
  y      <- gerabin(k,nis ,pis)
  res    <- MCTBinPar(k,nis ,y)
  etIePod[1,] <- etIePod[1,] + res$ctETI/nsim
  etIePod[2,] <- etIePod[2,] + res$ctPod/(nsim*k1*k2)
}
etIePod

# Avaliação do teste bayesiano(TB) para a igualdade das proporções
# binomiais computando o erro tipo I sob H0 completa e o poder sob
# H0 parcial.

k      <- 2
Delta  <- 0.0
nkisg1 <- 2
pi1    <- 0.01
# hiperparâmetros
alphas <- 2
alphas <- rep(c(alphas),k)
betas  <- 2
betas  <- rep(c(betas),k)
alpha  <- 2
beta   <- 2
# fim hiperparâmetros
nis    <- rep(c(10),k)
pis    <- c(rep(c(pi1),nkisg1),rep(c(pi1+Delta),k-nkisg1))
nsim   <- 1000
nsr = matrix(0,1,2)
rownames(nsr) = c("BTBin")
colnames(nsr)=c("alpha=0.05","alpha=0.01")
alph   <- c(0.05,0.01) # níveis significância
for (i in 1:nsim)
{
  y      <- gerabin(k,nis ,pis)

```

```

    qalpha <- quantisdeqj(k,nis,y,alpha,beta,alph,10000)
    chiB <- BTBin(k,nis,y,alphas,betas,alpha,beta,qalpha)
    nsr[chiB$sig==TRUE] <- nsr[chiB$sig==TRUE] + 1/nsim
  }
  nsr

# Avaliação do teste de comparações múltiplas bayesiano (TCMB)
# especificamente para computar o erro tipo I por experimento
# sob H0 completa

k <- 2
Delta <- 0.00
nkisg1 <- 2
pi1 <- 0.01
# hiperparâmetros
alphas <- 0.01
alphas <- rep(c(alphas),k)
betas <- 0.01
betas <- rep(c(betas),k)
alpha <- 2
beta <- 2
# fim hiperparâmetros
nis <- rep(c(10),k)
pis <- c(rep(c(pi1),nkisg1),rep(c(pi1+Delta),k-nkisg1))
nsim <- 1000
etIe = matrix(0,1,2)
rownames(etIe) = c("Erro Tipo I Por Experimento")
colnames(etIe)=c("alpha=0.05","alpha=0.01")
alph <- c(0.05,0.01) # níveis significância
for (i in 1:nsim)
{
  y <- gerabin(k,nis,pis)
  qc <- MCTBinBayes(k,nis,y,alphas,betas,alpha,beta,alph)
  ct <- MCTETIEBinBayes(qc)
  etIe[1,] <- etIe[1,] + ct/nsim
}
etIe

# Avaliação do teste de comparações múltiplas bayesiano especifi-
# camente para computar o erro tipo I por experimento e o poder
# sob H0 parcial.

k <- 5
pi1 <- 0.01
Delta <- 0.01
k1 <- k %/% 2
if (k %/% 2 != 0) k1 <- k1 + 1
k2 <- k - k1
# hiperparâmetros
alphas <- 0.01

```

```

alphas <- rep(c(alphas),k)
betas <- 0.01
betas <- rep(c(betas),k)
alpha <- 2
beta <- 2
# fim hiperparâmetros
nis <- rep(c(10),k)
pis <- c(rep(c(pi1),k1),rep(c(pi1+Delta),k2))
nsim <- 1000
etIePod = matrix(0,2,2)
rownames(etIePod) = c("ETI Por Experimento - H0 parcial",
" Poder - H0 parcial")
colnames(etIePod)=c("alpha=0.05","alpha=0.01")
alph <- c(0.05,0.01) # níveis significância
for (i in 1:nsim)
{
y <- gerabin(k,nis,pis)
res <- MCTBinParBayes(k,nis,y,alphas,betas,alpha,beta,alph)
etIePod[1,] <- etIePod[1,] + res$ctETI/nsim
etIePod[2,] <- etIePod[2,] + res$ctPod/(nsim*k1*k2)
}
etIePod

```

PROGRAMA B: Programa R para realizar inferências sobre proporções binomiais

```
# Programa para aplicação dos testes assintóticos de proporções
# binomiais e de comparações múltiplas utilizando inferências
# frequentistas e bayesianas.

# Função que retorna a média e a matriz de covariâncias da
# posteriori sob H1, referente as expressões (4.4) e (4.5) da tese,
# recebendo: o vetor y (k x 1) de observações, o vetor nis(k x 1)
# dos tamanhos amostrais, o número k de populações, o vetor alphas
# de hiperparâmetros alpha_is e o vetor betas dos hiperparâmetros
# beta_is (k x 1).

MedCovPostH1 <- function(k, nis, y, alphas, betas)
{
  pip <- (alphas+y)/(alphas+betas+nis)
  sigp <- (alphas+y)*(betas+nis-y)/((alphas+betas+nis)^2
    *(alphas+betas+nis+1))
  return(list(pip=pip, sigp=diag(sigp)))
}

# Função que retorna a média e a matriz de covariâncias da
# posteriori sob H0, referente as expressões (4.9) e (4.10)
# da tese, recebendo: o vetor y (k x 1) de observações, o
# vetor nis (k x 1) dos tamanhos amostrais, o número k de
# populações, o hiperparâmetro alpha e o hiperparâmetro beta.

MedCovPostH0 <- function(k, nis, y, alpha, beta)
{
  pi0 <- rep((alpha+sum(y))/(alpha+beta+sum(nis)), times=k)
  sig0 <- (alpha+sum(y))*(beta+sum(nis)-sum(y))/
    ((alpha+beta+sum(nis))^2*(alpha+beta+sum(nis)+1))
  sig0 <- rep(sig0, times=k)
  return(list(pi0=pi0, sig0=diag(sig0)))
}

# Função para obter a distribuição das formas quadráticas q_j,
# representada pela equação (4.18) da tese. Deve receber
# os hiperparâmetros alpha e beta, os tamanhos amostrais nis,
# as observações y (k x 1) e o número de populações k.

quantisdeqj <- function(k, nis, y, alpha, beta, alph, N=50000)
{
  pi0cov0 <- MedCovPostH0(k, nis, y, alpha, beta)
  alphap0 <- alpha + sum(y)
  betap0 <- beta + sum(nis) - sum(y)
  pi0p <- (matrix(rbeta(k*N, alphap0, betap0), k, N)
    -pi0cov0$pi0)^2
  pi0p <- pi0p/pi0cov0$sig0[1,1] # artifício
  pi0p <- sort(apply(pi0p, 2, sum))
}
```

```

qalpha      <- round((1-alpha)*N)
qalpha[qalpha==0] <- 1
qalpha      <- pi0p[qalpha]
return(qalpha)
}

# Função para aplicar o teste bayesiano (TB) com hiperparâmetros
# alphas, betas, alpha e beta.

BTBin      <- function(k, nis, y, alphas, betas, alpha, beta, qalpha)
{
  pi0cov0   <- MedCovPostH0(k, nis, y, alpha, beta)
  pi1cov1   <- MedCovPostH1(k, nis, y, alphas, betas)
  sig0inv   <- diag(1/diag(pi1cov1$sigp))
  qc        <- t(pi1cov1$pi1-pi0cov0$pi0)**%*%sig0inv**%*%
              (pi1cov1$pi1-pi0cov0$pi0)
  sig       <- rep(qc, times=length(qalpha))>qalpha
              # TRUE para significativo
  return(list(qc=qc))
}

# Função para aplicar o teste de razão de verossimilhanças (G2)
# da equação (4.11) da tese.

LRTBin     <- function(k, nis, y)
{
  hatpis    <- y/nis
  n         <- sum(nis)
  somaYj    <- sum(y)
  hatpi0    <- sum(y)/n
  aux       <- any(y>0)
  if (aux==FALSE) aux1 <- n*y else
  aux1      <- n*y/(nis*somaYj)
  aux       <- aux1>0
  aux1      <- sum(y[aux]*log(aux1[aux]))
  aux2      <- (nis-y)
  aux       <- aux2>0
  aux2      <- sum(aux2[aux]*log(n*aux2[aux]/(nis[aux]*(n-somaYj))))
  chi1      <- 2*(aux1+aux2)
  df        <- k -1
  pr.chi    <- pchisq(chi1, df, lower.tail=FALSE)
  return(list(hatpis=hatpis, hatpi0=hatpi0, chi=chi1, df=df,
              pr.chi=pr.chi))
}

# Função para aplicar o teste de formas quadráticas (X2) da
# equação (4.13) da tese.

QFTBin     <- function(k, nis, y)
{

```

```

    hatpi0 <- sum(y)/sum(nis)
    aux1   <- (y-nis*hatpi0)^2
    aux2   <- nis*hatpi0*(1-hatpi0)
    aux    <- !any(aux2<=0)
    if (aux==TRUE) chi2 <- sum(aux1/aux2) else chi2 <- 0
    df     <- k - 1
    pr.chi <- pchisq(chi2,df,lower.tail=FALSE)
    return(list(hatpi0=hatpi0,chi=chi2,df=df,pr.chi=pr.chi))
  }

# Função para aplicar o teste de comparações múltiplas (TCM),
# representado pela equação (5.2) da tese.

MCTBin <- function(k,nis,y)
{
  res <- diag(y/nis)
  hatpi0 <- sum(y)/sum(nis)
  varpi <- hatpi0*(1-hatpi0)/nis
  df <- k - 1
  for (i in 1:(k-1))
  {
    for (j in (i+1):k)
    {
      if ((varpi[i]==0) & (varpi[j]==0))
      {
        res[i,j] <- 0
      } else
        res[i,j] <- (res[i,i]-res[j,j])^2/(varpi[i]+varpi[j])
        res[j,i] <- pchisq(res[i,j],df,lower.tail=FALSE)
    }
  }
  return(res)
}

# Função para aplicar o teste de comparações múltiplas Bayesiano
# (TCMB) da tese.

MCTBinBayes <- function(k,nis,y,alphas,betas,alpha,beta,
                        alph=c(0.05,0.01))
{
  pilcov1 <- MedCovPostH1(k,nis,y,alphas,betas)
  qalpha <- quantisdeqj(k,nis,y,alpha,beta,alph,10000)
  qalpha1 <- matrix(0,k,k)
  qalpha2 <- matrix(0,k,k)
  for (i in 1:(k-1))
  {
    elle <- matrix(0,k,1)
    elle[i] <- 1
    for (j in (i+1):k)
    {

```

```

    elle[j]      <- -1
    if ((pilcov1$sigp[i,i]==0) & (pilcov1$sigp[j,j]==0))
    {
      qalpha1[i,j] <- 0
      qalpha2[i,j] <- 0
    } else
    {
      aux          <- pilcov1$sigp[i,i]+pilcov1$sigp[j,j]
      qalpha1[i,j] <- (t(elle)%*%pilcov1$pip)/sqrt(aux)
      qalpha2[i,j] <- qalpha1[i,j]
      if (abs(qalpha1[i,j]) >= sqrt(qalpha[1]))
      qalpha1[j,i] <- TRUE else qalpha1[j,i] <- FALSE
      if (abs(qalpha2[i,j]) >= sqrt(qalpha[2]))
      qalpha2[j,i] <- TRUE else qalpha2[j,i] <- FALSE
    }
    elle[j] <- 0
  }
}
return(list(qc1=qalpha1 ,qc2=qalpha2))
}

# Dados a serem fornecidos para realização de todos os testes

k      <- 5
y      <- c(20, 18, 25, 29, 30)
nis    <- c(30, 30, 30, 30, 30)

# Dados a serem fornecidos para realização dos testes bayesianos

# hiperparâmetros
alphas <- 0.01
alphas <- rep(c(alphas),k)
betas  <- 0.01
betas  <- rep(c(betas),k)
alpha  <- 2
beta   <- 2
# fim hiperparâmetros

# Resultados dos testes G2 e X2:

# G2 retorna: os valores dos estimadores de todos os pis;
# o valor do estimador de pi comum; a estatística do teste;
# os graus de liberdade e o valor-p referente à estatística.

# X2 retorna: o valor do estimador de pi comum; a estatística
# do teste; os graus de liberdade e o valor-p referente à
# estatística.

G2 <- LRTBin(k,nis,y)
X2 <- QFTBin(k,nis,y)

```

```

G2
X2

# Resultados do teste de comparações múltiplas (TCM):

# TCM retorna na diagonal os valores dos estimadores dos pi
# Acima da diagonal os valores da estatística (pi vs pj)
# Abaixo da diagonal os valores-p referente à estatística

TCM <- MCTBin(k,nis,y)
TCM

# Resultado do teste bayesiano (TB):

# qalpha retorna os quantis da distribuição nula a 5 e 1%,
# respectivamente.
# TB retorna o valor da estatística

qalpha <- quantisdeqj(k,nis,y,alpha,beta,alpha=c(0.05,0.01),10000)
TB <- BTBin(k,nis,y,alphas,betas,alpha,beta,qalpha)
qalpha
TB

# Resultado do teste de comparações múltiplas bayesiano (TCMB)
# das proporções binomiais

# qc1 retorna as inferências realizadas para o nível nominal
# de 5%. Sendo que:
# Na diagonal tem-se os valores das médias da posteriori
# de pi sob H1
# Acima da diagonal tem-se os valores da estatística qc (pi vs pj)
# Abaixo da diagonal tem-se houve ou significância entre as
# proporções binomiais, recebendo 1 se significativo e 0 se
# não significativo.

# qc2 retorna as inferências realizadas para o nível nominal
# de 1%. Idem qc1

qc <- MCTBinBayes(k,nis,y,alphas,betas,alpha,beta,
  alpha=c(0.05,0.01))
qc

```