Comparações múltiplas para dados censurados

Daiane de Souza Santos

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: _____

Comparações múltiplas para dados censurados

Daiane de Souza Santos

***Orientador:* Prof. Dr. Dorival Leão Pinto Júnior**

Dissertação apresentada ao Instituto de Ciências Matemáticas e de Computação - ICMC-USP, como parte dos requisitos para obtenção do título de Mestre em Ciências – Ciências de Computação e Matemática Computacional. *VERSÃO REVISADA*

USP – São Carlos
Junho de 2013

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados fornecidos pelo(a) autor(a)

d237c de Souza Santos, Daiane
Comparações múltiplas para dados censurados /
Daiane de Souza Santos; orientador Dorival Leão
Pinto Júnior. -- São Carlos, 2013.
71 p.

Tese (Doutorado - Programa de Pós-Graduação em
Ciências de Computação e Matemática Computacional) --
Instituto de Ciências Matemáticas e de Computação,
Universidade de São Paulo, 2013.

1. Comparações múltiplas. I. Leão Pinto Júnior,
Dorival , orient. II. Título.

Comparações múltiplas para dados censurados

Daiane de Souza Santos

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: _____

Comparações múltiplas para dados censurados

Daiane de Souza Santos

***Orientador:* Prof. Dr. Dorival Leão Pinto Júnior**

Dissertação apresentada ao Instituto de Ciências Matemáticas e de Computação - ICMC-USP, como parte dos requisitos para obtenção do título de Mestre em Ciências – Ciências de Computação e Matemática Computacional. *VERSÃO REVISADA*

USP – São Carlos
Junho de 2013

Agradecimentos

Primeiramente, agradeço à minha mãe Vilma, à minha avó Olita e ao meu tio Alexandre por sempre me apoiarem e incentivarem.

Ao meu orientador, Dorival Leão, por ter acreditado em mim e por estar presente em cada etapa deste trabalho. Agradeço ainda pela paciência e incentivo na elaboração e condução deste trabalho. Minha eterna gratidão e admiração.

Aos professores de estatística do ICMC-USP pelos ensinamentos e por toda atenção que me deram. Em especial à professora Cibele Russo, membro da banca do exame de qualificação, agradeço pelas diretrizes e sugestões.

Aos professores e funcionários do ICMC pela minha formação profissional, pessoal e pelo excelente convívio.

Em especial ao Roberto pelo apoio e carinho, ao Bruno e à Naiara pelos preciosos momentos de distração e à Patrícia, que mesmo à distância está todo dia presente.

A todos meus amigos do curso de pós graduação em estatística: Alessandra, Aline, Hiroshi, Gilberto, João Luiz, José Augusto, José Flores Delgado, Matheus e Willian, por toda ajuda que me deram e pelos momentos compartilhados ao longo destes anos.

Às minhas queridas companheiras de casa Nayara e Ludmila e aos demais amigos que, independentemente da distância acompanham-me nesta jornada. Não citarei nomes, pois esta lista seria muito extensa para ser colocada aqui. Além disso, não correrei o risco de acabar esquecendo de alguém.

A todas as pessoas que não foram mencionadas, mas que de alguma maneira contribuíram para viabilizar este trabalho.

Resumo

O objetivo deste trabalho é estudar a performance de alguns métodos de comparações múltiplas (MCMs) que ajustam o valor-p quando as estatísticas empregadas nos testes são a *log-rank* e a Cramér-von Mises, ambas não paramétricas e com estrutura de dependência. A vantagem dos MCMs que ajustam o valor-p é que eles controlam as taxas de erro tipo I e tipo II para cada hipótese, a fim de atingir um poder estatístico elevado, mantendo a taxa de erro da família dos testes (FWER) menor ou igual ao nível de significância escolhido. Trabalhamos com o procedimento clássico de Bonferroni e com outros métodos vistos como seu melhoramento, com especial atenção a certos procedimentos derivados do método de Simes que permitem realizar inferências sob as hipóteses individuais. Foi verificado teoricamente que a estatística *log-rank* pertence à classe multivariada totalmente positiva de ordem 2 (MTP_2), uma vez que o método de Simes garante o controle da FWER quando as estatísticas dependentes assumem esta condição. O controle da FWER empregando a estatística de Cramér-von Mises foi observado apenas por meio de simulações. Os MCMs foram analisados através de estudos computacionais em modelos discretos e contínuos sob censura com foco no problema de comparar um tratamento versus controle.

Palavras-chave: Métodos de comparações múltiplas, controle da FWER, melhoramentos do método de Bonferroni, método de Simes, estatísticas dependentes, dados censurados.

Abstract

The aim of this work is to study the performance of some Multiple Comparison Methods (MCMs) that adjust the p-value when the log-rank-type and Cramér-von Mises statistics are used, both nonparametric and with dependency structure. The advantage of these methods is that they control the error rates of type I and type II for each hypothesis in order to achieve high statistical power while keeping the Family Wise Error Rate (FWER) lower or equal than a given significance level. The classical Bonferroni procedure is used as well as others seen as its improvement, with special attention to certain procedures derived from Simes' method for making inferences on individual hypothesis. It is theoretically proved that the weighted Log-Rank statistics belongs to the multivariate totally positive of order 2 (MTP_2) class, which is needed in order to apply Simes' method, that guarantees control of the FWER of dependent statistics in this case. The control of the FWER when the Cramér-von Mises statistics is used is only verified by means of computational simulations. The MCMs are also analyzed by means of computational experiments with discrete and continuous data under censoring with focus on the problem of comparisons of treatment versus a control.

Keywords: Multiple comparison methods, FWER control, improved Bonferroni method, Simes' method, dependent statistics, censoring data.

Sumário

1	Introdução	1
1.1	Organização dos capítulos	5
2	Testes Múltiplos e MCMS	7
2.1	Introdução	7
2.2	Conceitos principais	9
2.2.1	Taxa de erros da família dos testes - FWER	9
2.2.2	Valor descritivo do teste	9
2.2.3	Procedimentos <i>stepup</i> e <i>stepdown</i>	12
2.2.4	Métodos fechados	12
2.3	Procedimentos de Bonferroni	14
2.3.1	Bonferroni Clássico	14
2.3.2	Procedimento de Bonferroni de rejeição sequencial	16
2.4	Método de Simes	17
2.5	Método de Simes para testes individuais	22
2.5.1	Método de Hochberg	23
2.5.2	Método de Hommel	23
2.5.3	Método de Rom	25

3	Modelo e Estatísticas dos Testes	26
3.1	Introdução	26
3.2	Modelo para os dados	27
3.3	Estatísticas dos Testes	29
3.3.1	Estatística <i>log-rank</i>	29
3.3.2	Condição MTP_2 para a estatística <i>log-Rank</i>	31
3.3.3	Estatística de Cramér-von Mises	34
4	Estudo de simulação	37
4.1	Dados discretos	38
4.1.1	Controle da FWER	38
4.1.2	Poder dos testes	42
4.2	Dados contínuos	53
4.2.1	Controle da FWER	53
4.2.2	Poder dos testes	55
5	Conclusão e propostas futuras	61
A	Propriedades e resultados assintóticos	63
	Referências	70

Capítulo 1

Introdução

Testes simultâneos de várias hipóteses nulas, conhecidos simplesmente como testes múltiplos, são um dos problemas mais comuns de inferência em diversas análises estatísticas. Por exemplo, em pesquisas farmacêuticas sobre a eficácia de uma droga experimental, a avaliação estatística sobre a superioridade da droga em relação aos medicamentos existentes ou sobre o placebo é geralmente realizada por meio de uma formulação de testes múltiplos.

O principal desafio dos testes múltiplos está em assegurar que a taxa de erro Tipo I, que é rejeitar erroneamente a hipótese nula e, por conseguinte, chegar a uma inferência falso positiva, seja mantido ao nível estabelecido pelo investigador. Pesquisas biomédicas, por exemplo, geralmente utilizam projetos experimentais complexos. Dessa forma, a fim de economizar recursos, como dinheiro, tempo do investigador ou unidades experimentais (humanos, animais, tecidos ou células), várias hipóteses acabam sendo testadas a partir de uma única experiência, o que aumenta o risco de fazer inferências estatísticas falsas. No pior dos casos, se m hipóteses independentes são testadas, o risco de uma inferência errônea aumenta m vezes. É neste contexto que surgem os Métodos de Comparações Múltiplas (MCMs), que são procedimentos estatísticos designados a minimizar estes erros. A bibliografia sobre este assunto é extensa, podemos destacar como referências fundamentais os livros Hochberg & Tamhane (1987), Hsu (1996) e Lehmann & Romano (2005), que trazem quase todas as técnicas e conceitos empregados nesta área da inferência múltipla.

Na literatura de testes múltiplos diferentes taxas de erros foram analisadas, sendo as principais denominadas como: taxa de erro por comparação (*per-comparison error rate* (PCE)), definida como a proporção esperada do erro tipo I; taxa de erro por família (*error rate per family* (PFE)), exibida como o número de erros esperados na família; taxa de erros da família dos testes (*family-wise error rate* (FWER)), que representa a probabilidade de uma ou mais rejeições falsas na família de inferências; e ainda a *false discovery rate* (FDR), que é indicada quando temos um número bem elevado de hipóteses a serem testadas, e é definida pelo valor esperado da razão entre o número de rejeições falsas e o número total de rejeições. Neste trabalho, as discussões serão voltadas para os MCMs que controlam a FWER, que entre outras razões, segundo Hochberg & Tamhane (1987), nos permite calibrar uniformemente procedimentos diferentes para um referencial comum, e assim, comparar as suas características operacionais de maneira justa.

Um dos precursores dos MCMs foi Fisher (1935), que introduziu duas categorias distintas: procedimentos de uma etapa (*single-step*), que abrange a importante classe referida como procedimentos de testes simultâneos (*simultaneous test procedures*); e procedimentos de etapas múltiplas (*stepwise*), que posteriormente foram classificadas em “ascendente” (*stepup*) e “descendente” (*stepdown*).

O primeiro método proposto por Fisher é o famoso teste da diferença mínima significativa (*least significant difference* (LSD) test), que pode ser visto como um procedimento de duas etapas, em que a hipótese nula é testada preliminarmente por um teste F . Já o segundo método dado por Fisher apresenta etapa única e é popularmente conhecido como procedimento de Bonferroni. Ambos procedimentos controlam a FWER, o LSD no sentido fraco (somente quando todas as hipóteses nulas na família são verdadeiras) e Bonferroni no sentido forte (para qualquer combinação de hipóteses verdadeiras e falsas). No entanto, frequentemente o LSD exhibe maior poder e uma das razões disto é o fato de apresentar mais de uma etapa.

As comparações múltiplas na análise de variância (ANOVA) é uma área em que as inferências simultâneas têm uma longa história. O procedimento clássico e mais utilizado neste contexto foi proposto por Tukey (1953). Tal método compara todos os possíveis pares de médias a fim de encontrar quais delas diferem significativamente uma das outras, e é baseado na distribuição da amplitude Studentizada. Para dados balanceados, o teste de

Tukey é o método de uma etapa mais poderoso para comparações de médias duas a duas. Quando o interesse está na comparação de vários tratamentos com um controle, o teste que merece destaque na ANOVA foi proposto por Dunnett (1955), e é bem semelhante ao teste t usual.

Marcus *et al.* (1976) propuseram uma técnica geral para a construção de procedimentos descendentes: os métodos fechados (*closure method*). Os procedimentos resultantes deste método são a base para a construção de muitas estratégias de comparações que veremos neste trabalho, eles controlam a FWER no sentido forte e é chamado na literatura de procedimentos de testes fechados (*closed testing procedures*).

O método de Bonferroni clássico tem presença marcante na inferência simultânea, pode ser aplicado em qualquer situação de testes múltiplos e como dito anteriormente, controla a FWER no sentido forte, porém o preço desta generalidade é a falta de poder. Diante disto, a fim de amenizar esta situação, tem surgido na literatura diversos procedimentos denominados como “melhoramentos de Bonferroni”, já que um MCM ideal é aquele que controla a FWER e tem capacidade para detectar casos em que a hipótese nula é falsa. Estas técnicas geralmente apresentam etapas múltiplas e são propostas para ajustar o valor-p oriundo das estatísticas dos testes. A vantagem dos MCMs que ajustam o valor-p é que eles controlam as taxas de erro tipo I e tipo II para cada hipótese com o propósito de atingir um poder estatístico elevado, mantendo a FWER menor ou igual ao nível de significância escolhido. Um dos primeiros aprimoramentos do método clássico foi dado por Holm (1979), que propôs um método descendente através do procedimento dos testes fechados, que ficou conhecido como procedimento de Bonferroni de rejeição sequencial (*sequentially rejective bonferroni procedure*) ou procedimento de Holm. Esta estratégia exibe maior poder quando comparado ao método tradicional de Bonferroni e controla a FWER no sentido forte.

Uma das modificações mais importantes do procedimento de Bonferroni e que daremos grande atenção neste trabalho foi dado por Simes (1986). Tal método testa a hipótese nula global $H_0 = \bigcap_{i=1}^n H_i$, que é a combinação de n hipóteses individuais, a um nível de significância múltiplo α pré-estabelecido em termos dos valores-p das estatísticas dos testes. Simes provou que seu procedimento controla a FWER quando as estatísticas dos testes são independentes e conjecturou via simulação que este controle se mantém

para algumas variedades de estatísticas multivariadas dependentes. Evidenciamos este método pelo fato dele apresentar maior poder em relação a Bonferroni, principalmente quando estatísticas altamente correlacionadas estão envolvidas.

Como Simes (1986) não justificou teoricamente o controle da FWER para estatísticas dependentes, Sarkar & Chang (1997) mostraram que para distribuições multivariadas que exibem um certo tipo de dependência positiva o método de Simes controla efetivamente a FWER. Posteriormente, Sarkar (1998) provou analiticamente que o método de Simes controla a FWER para uma classe ainda mais geral de distribuições multivariadas com dependência positiva, caracterizadas pela condição multivariada totalmente positiva de ordem 2 (MTP_2) e também para algumas misturas de MTP_2 .

Simes (1986) também deixou em aberto o problema de realizar inferências simultâneas sob as hipóteses individuais. Diante disto, Hochberg (1988) sugeriu um método bem simples de rejeição sequencial do tipo ascendente. Hommel (1988) também estendeu o procedimento de Simes para inferências individuais, e subsequentemente, Hommel (1989) mostrou que este procedimento, embora seja mais complicado e difícil de ser interpretado, é mais poderoso que a estratégia formulada por Hochberg. A fim de melhorar o procedimento de Hochberg, Rom (1990) propôs uma modificação que requer o cálculo dos pontos críticos através de um algoritmo recursivo, e foi mostrado que tal procedimento apresenta uma pequena vantagem em relação ao método original. Recentemente, Rom (2013) forneceu uma modificação que mantém a simplicidade do procedimento de Hochberg enquanto apresenta um ganho de poder, que pode ser moderado em algumas situações, podendo atingir níveis bem elevados quando muitas das hipóteses testadas são falsas.

Neste trabalho o principal objetivo é aplicar os procedimentos de Bonferroni, Holm, Hochberg, Hommel e o método recente proposto por Rom (2013) em modelos discretos sob censura aleatória, já que a análise estatística para estes dados tem grande importância em inúmeras áreas. Um exemplo típico visto na educação é quando queremos analisar o número de semestres até a ocorrência da evasão dos alunos de graduação de uma universidade. Quando analisamos dados ao longo do tempo como estes, um problema frequente é que alguns dados aparecem somente em certos períodos, ou seja, estão sujeitos à censura. Dessa forma, quando o evento de interesse é a ocorrência da evasão devemos considerar os eventos de censura, que neste caso podem ser a transferência, a migração,

o falecimento ou a conclusão do curso.

Em várias circunstâncias, procedimentos estatísticos baseados em distribuições contínuas ou mistas não podem ser aplicados diretamente em dados puramente discretos, devido à falta de informação necessária para validar os resultados assintóticos. Um problema encontrado na literatura está no desenvolvimento de métodos de testes não paramétricos para modelos discretos na presença de censura que seja consistente em uma grande classe de hipóteses alternativas. Um dos métodos não paramétricos mais importantes para testar H_0 sob censura aleatória é o teste de *log-rank*. De fato, é um dos testes não paramétrico mais indicados para comparar duas ou mais populações com dados sujeitos à censura, contudo apresenta a desvantagem de não exibir um bom poder ao lidar com funções intensidade não proporcionais. Dessa maneira, trabalhamos também com a estatística de Cramér-von Mises, que foi analisada em Leão & Ohashi (2011) e considerada ideal para modelos puramente discretos sob censura.

Verificamos teoricamente que a estatística *log-rank* satisfaz a condição MTP_2 . O controle da FWER ao empregar as extensões do método de Simes na estatística Cramér-von Mises foi averiguado apenas por recursos computacionais. Ainda por meio de estudos de simulações comparamos a performance dos métodos de Bonferroni, Holm, Hochberg, Hommel e Rom em relação ao ganho de poder realizando os testes com as estatísticas *log-rank* e Cramér-von Mises. Priorizamos nos estudos dados gerados de distribuição discreta, mas também realizamos uma pequena análise com dados contínuos. Focamos no problema de comparar um tratamento versus controle.

1.1 Organização dos capítulos

O foco do Capítulo 2 é destacar os principais conceitos empregados na teoria de comparações múltiplas e apresentar detalhadamente o procedimento de Simes e suas principais extensões para testes individuais.

No Capítulo 3 introduzimos o modelo discreto sob censura aleatória com o qual trabalhamos e apresentamos as estatísticas dos testes empregadas: *log-rank* e Cramér-von Mises. Mostramos ainda que a estatística *log-rank* satisfaz a condição MTP_2 para a aplicação do método de Simes.

O Capítulo 4 traz um estudo de simulação que realizamos utilizando a linguagem R (R Development Core Team, 2011), com o objetivo de analisar o desempenho do método de Simes e suas extensões para testes individuais, quanto ao controle da FWER e ao ganho de poder ao realizar os testes com as estatísticas *log-rank* e Cramér-von Mises. Por fim, o capítulo 5 apresenta as conclusões dos resultados obtidos e propostas de trabalhos futuros.

Capítulo 2

Testes Múltiplos e MCMs

Neste capítulo exploramos os conceitos mais importantes envolvidos na teoria de testes múltiplos, principalmente no que diz respeito aos MCMs que controlam a FWER no sentido forte e que são descritos por meio dos valores-p oriundos das estatísticas dos testes. Relatamos as modificações mais relevantes do método clássico de Bonferroni, apresentamos o método de Simes e algumas de suas extensões para realizar inferências sob as hipóteses nulas individuais.

2.1 Introdução

Em pesquisas médicas, educacionais ou agrícolas, uma ferramenta comum para a análise dos dados é um teste de homogeneidade das populações (tratamentos ou grupos); e quando a hipótese nula que carrega a informação de que todas as populações são iguais é rejeitada, não obtemos inferências sobre diversas comparações detalhadas entre as populações que podem ser de interesse do investigador, por exemplo, comparações dois a dois entre os grupos, comparações de vários tratamentos com um controle (ou referência), ou ainda comparações entre todos os subconjuntos dos grupos. É neste contexto que surgem os Testes Múltiplos, em especial as Comparações Múltiplas, já que nossa atenção neste trabalho está voltada às comparações.

Quando realizamos testes de hipóteses podemos cometer dois possíveis erros na tomada de decisão: o erro tipo I, que ocorre quando a hipótese nula H_0 é rejeitada

erroneamente e o erro tipo II, que ocorre ao não rejeitarmos H_0 quando esta é falsa.

É desejável realizar os testes de uma maneira que as probabilidades dos dois tipos de erro sejam mínimas, porém quando temos um número grande de hipóteses a serem testadas simultaneamente nos deparamos com o problema da multiplicidade, que se refere ao crescimento da probabilidade do erro tipo I com o aumento do número de testes. Para ilustrar numericamente este fato vejamos o exemplo a seguir, em que foi calculada a probabilidade de ao menos uma rejeição falsa quando todas as hipóteses $H_1, \dots, H_n$ são verdadeiras e as estatísticas dos testes são independentes.

Considerando o nível de significância $\alpha = 0,05$ para cada teste individual, temos que o nível de significância conjunto (NSC) é dado por

$$P\{\text{ao menos uma rejeição falsa}\} = 1 - \prod_{i=1}^n \{1 - P(\text{rejeição falsa})\} = 1 - (1 - \alpha)^n.$$

Dessa forma, se executarmos 10 testes de hipóteses, a probabilidade de rejeitarmos pelo menos uma hipótese verdadeira é 0,40; se forem executados 20 testes, a probabilidade aumenta para 0,62 e se forem realizados 50 testes, a probabilidade chega a 0,92. O gráfico na sequência mostra o comportamento do NSC à medida que aumentamos o número de testes realizados.

FIGURA 2.1: Comportamento do NSC.

É diante deste problema que os MCMs foram formulados, para levar em conta e controlar adequadamente o nível de significância conjunto, isto é, garantir que a probabilidade de ao menos uma rejeição falsa não aumente com o número de testes envolvidos.

2.2 Conceitos principais

2.2.1 Taxa de erros da família dos testes - FWER

Para proporcionarmos uma abordagem objetiva para o problema de inferências múltiplas, garantindo uma correção simultânea para um conjunto de inferências e chegar a uma decisão global correta, introduzimos o conceito de *família*, que pode ser definida como qualquer coleção de hipóteses que estão submetidas a testes comuns, em que os erros conjuntos são controlados. Por exemplo, num teste laboratorial de amostras de sangue, uma família pode ser definida como todos os ensaios realizados em um dia, ou aqueles realizados em um dia por um determinado testador. Por outro lado, os testes realizados pela manhã e tarde podem ser considerados como famílias distintas, e assim por diante. Em suma, a escolha dos testes que serão tratados como uma família dependerá da situação (Lehmann & Romano, 2005). Todos os MCMs discutidos neste trabalho são definidos em termos de uma família finita de hipóteses.

Como nosso enfoque não está na realização de um número muito extenso de testes e temos como objetivo chegar a uma decisão final satisfatória, é fundamental que todas as inferências simultâneas sejam corretas, assim é necessário que os MCMs controlem a taxa de erro da família de testes (*family-wise error rate*), que denotaremos a partir de agora como FWER. Esta taxa representa a probabilidade de cometermos ao menos uma rejeição falsa na família. Portanto, uma vez que a família de testes foi definida, queremos que

$$\text{FWER} \leq \alpha,$$

para todas as possíveis coleções de hipóteses verdadeiras e falsas.

2.2.2 Valor descritivo do teste

Sem dúvida o valor descritivo do teste, conhecido popularmente como valor-p, tornou-se o “ponto de partida” para muitos que utilizam análise estatística; e na inferência simultânea isto não é diferente, já que um valor-p fornece informações acerca se o teste de hipótese é significativo ou não, e ainda indica o “quanto significativo” o resultado é,

visto que, quanto menor o valor-p maiores as evidências contra a hipótese nula.

Testes estatísticos da intersecção de um número de hipóteses nulas com base em seus valores-p correspondentes são um problema comum na prática. Por exemplo, na avaliação da superioridade de uma droga nova em relação a uma droga existente ou placebo, estudos múltiplos são frequentemente utilizados para investigar diferentes aspectos dos efeitos das drogas. Transmitir resultados dos estudos individuais tornou-se uma prática comum na pesquisa farmacêutica. Assim, para obter uma conclusão, por exemplo, se existe ou não ao menos uma dose ou grupo de pacientes para os quais a nova droga é mais eficiente do que seu concorrente, é muitas vezes conveniente utilizar apenas os valores-p correspondentes às comparações individuais (Sarkar & Chang, 1997). Diante disto, na sequência relatamos algumas propriedades do valor-p que são fundamentais para a teoria dos MCMs que tratamos neste trabalho.

A tomada de decisão em um teste de hipótese é baseada no valor de uma certa variável aleatória X , com distribuição P_θ pertencente à classe $\mathcal{P} = \{P_\theta, \theta \in \Omega\}$, em que θ é o parâmetro de interesse e Ω é o espaço paramétrico. Definimos como H a classe das hipóteses nulas e K a classe das hipóteses alternativas, e os subconjuntos correspondentes de Ω por Ω_H e Ω_K respectivamente, tais que $H \cup K = \mathcal{P}$ e $\Omega_H \cup \Omega_K = \Omega$.

Um procedimento de teste não aleatorizado atribui a cada possível valor x que X pode assumir a decisão de rejeitar ou não rejeitar H , e assim, divide o espaço amostral em duas regiões complementares: região de aceitação e região de rejeição de H .

Consideremos S_α uma região de rejeição de nível α . Variando valores de α , estas regiões podem ser encaixadas no sentido que

$$S_\alpha \subset S_{\alpha'} \quad \text{se} \quad \alpha < \alpha'. \quad (2.1)$$

Sob essa situação, conseguimos determinar o menor nível de significância, valor-p, definido matematicamente por

$$p = p(X) = \inf \{ \alpha : X \in S_\alpha \}.$$

Lema 2.1 (Lehmann & Romano (2005)). *Suponhamos que X tenha distribuição de probabilidade P_θ para algum $\theta \in \Omega$, e a hipótese nula H estabelece $\theta \in \Omega_H$. Assumimos que as regiões de rejeição sejam encaixantes como em (2.1).*

(i) Se

$$\sup_{\theta \in \Omega_H} P_\theta \{X \in S_\alpha\} \leq \alpha, \text{ para todo } 0 < \alpha < 1, \quad (2.2)$$

então a distribuição de p sobre $\theta \in \Omega_H$ satisfaz

$$P_\theta \{p \leq u\} \leq u \text{ para todo } 0 \leq u \leq 1.$$

(ii) Se, para $\theta \in \Omega_H$,

$$P_\theta \{X \in S_\alpha\} = \alpha \text{ para todo } 0 < \alpha < 1, \quad (2.3)$$

então,

$$P_\theta \{p \leq u\} = u \text{ para todo } 0 \leq u \leq 1;$$

ou seja, p tem distribuição uniforme $(0, 1)$.

Demonstração. (i) Se $\theta \in \Omega_H$, pela definição de valor-p, para todo $v > u$, $\{p \leq u\} \subset \{X \in S_v\}$, o que implica em $P_\theta \{p \leq u\} \leq P_\theta \{X \in S_v\}$. Dessa forma,

$$\lim_{v \rightarrow u^+} P_\theta \{p \leq u\} \leq \lim_{v \rightarrow u^+} P_\theta \{X \in S_v\},$$

e desde que vale (2.2), temos que $P_\theta \{p \leq u\} \leq u$.

(ii) Novamente pela definição do valor-p, desde que o evento $\{X \in S_u\}$ implica que $\{p \leq u\}$, temos

$$P_\theta \{p \leq u\} \geq P_\theta \{X \in S_u\}.$$

Logo, desde que (2.3) é mantida, $P_\theta \{p \leq u\} \geq u$. Portanto, pelo resultado obtido em (i) concluímos que $P_\theta \{p \leq u\} = u$. $\square$

A maioria dos MCMs que controlam a FWER no sentido forte são denotados a partir dos valores-p, as hipóteses geralmente são ordenadas de acordo com seus níveis descritivos e então estes valores são ajustados (corrigidos). O valor-p ajustado para uma hipótese específica dentro de um conjunto de hipóteses é o menor nível de significância global em que a hipótese particular seria rejeitada. Um valor-p ajustado poder ser comparado diretamente com qualquer nível de significância α escolhido: se o valor-p ajustado for menor ou igual a α , a hipótese será rejeitada (Wright, 1992).

2.2.3 Procedimentos *stepup* e *stepdown*

Os MCMs de etapas múltiplas (*stepwise*) podem ser classificados em ascendentes (*stepup*) ou descendentes (*stepdown*). Um procedimento descendente é iniciado determinando se o teste que parece ser mais significativo pode ser rejeitado ou não. Se cada teste individual representado por um valor-p, o método descendente é descrito da seguinte maneira. Consideramos as constantes

$$\alpha_1 \leq \alpha_2 \leq \dots \leq \alpha_s. \quad (2.4)$$

Sejam $p_{(1)}, \dots, p_{(r)}$, a ordenação de $p_1, \dots, p_r$. Se $p_{(1)} \geq \alpha_1$, nenhuma hipótese é rejeitada. Caso contrário, para $r = 1, \dots, n$, as hipóteses $H_{(1)}, \dots, H_{(r)}$, associadas a $p_1, \dots, p_r$, são rejeitadas se,

$$p_{(1)} < \alpha_1, \dots, p_{(r)} < \alpha_r.$$

Em suma, o procedimento descendente inicia-se com o valor-p mais significativo e segue rejeitando as hipóteses contanto que seus valores-p sejam pequenos.

Em contrapartida, um procedimento ascendente inicia-se avaliando o valor-p menos significativo. Para o conjunto de constantes dado em (2.4), todas as hipóteses são rejeitadas se $p_{(n)} < \alpha_n$. Caso contrário, para $r = n, n-1, \dots, 1$, as hipóteses $H_{(1)}, \dots, H_{(r)}$ são rejeitadas se

$$p_{(n)} \geq \alpha_n, \dots, p_{(r+1)} \geq \alpha_{r+1} \quad \text{mas} \quad p_{(r)} < \alpha_r.$$

Os métodos de Holm e de Hochberg que veremos adiante são exemplos de procedimentos descendentes e ascendentes, respectivamente.

2.2.4 Métodos fechados

O método proposto por Holm (1979), bem como o método ainda mais poderoso dado por Hommel (1988), que veremos na sequência deste capítulo, são baseados no princípio dos testes fechados de Marcus *et al.* (1976), que segundo Hochberg & Tamhane (1987) podem ser vistos como métodos gerais para construção de procedimentos de testes descendentes. A versão deste procedimento que daremos a seguir pode ser vista em

Hommel (1986, 1988).

Seja $H_1, \dots, H_n$ uma coleção de n hipóteses. Definimos todas as possíveis combinações dos subconjuntos destas hipóteses como $H_I = \cap\{H_i : i \in I\}$, para todo $I \in \mathcal{Q}$, em que $\mathcal{Q}$ é o conjunto de todos subconjuntos não vazios de $\{1, 2, \dots, n\}$. Suponhamos que para cada H_I existe um teste baseado numa estatística T_I . Para um dado α , H_I é rejeitado se cada H_J é rejeitado ao nível α pela estatística T_J correspondente, em que $J \in \mathcal{Q}$ e $J \supseteq I$. A probabilidade de uma ou mais rejeições falsas quando testamos todas as hipóteses H_I é menor ou igual a α , como mostra o próximo teorema.

Teorema 2.1 (Marcus *et al.* (1976)). *O procedimento de testes fechados (PTF) controla a FWER no sentido forte ao nível α .*

Demonstração. Seja $\{H_i, i \in I\}$ uma coleção qualquer de hipóteses nulas verdadeiras e seja $H_I = \cap\{H_i : i \in I\}$, em que I é algum subconjunto desconhecido de $\{1, 2, \dots, n\}$. Se I é vazio, não pode haver erro tipo I, então consideramos $I \neq \emptyset$. Definimos A como o evento em que ao menos uma hipótese verdadeira H_i é rejeitada e B o evento em que H_I é rejeitada.

O procedimento de testes fechados rejeita uma hipótese H_i verdadeira se, e somente se, todas as hipóteses que implicam H_i , em particular H_I , são testadas ao nível α e são rejeitadas; e desde que o teste de H_i também seja significativo ao nível α . Assim, podemos tomar $A = A \cap B$, então sob H_I

$$\text{FWER} = P\{A\} = P\{A \cap B\} = P\{B\}P\{A|B\} \leq \alpha. \quad (2.5)$$

A última desigualdade segue desde que $P\{B\} \leq \alpha$ quando H_I é verdadeira. Como a desigualdade (2.5) se mantém em qualquer H_I , o teorema está provado. $\square$

Para aplicar o PTF podemos inicialmente começar com o teste global $H_I = \cap\{H_i : i = 1, 2, \dots, n\}$. Se este teste é rejeitado ao nível α , o procedimento de teste ainda continua ao nível α para cada subconjunto com $n - 1$ hipóteses. Enquanto as hipóteses continuam sendo rejeitadas ao nível α , o teste continua até atingir subconjuntos de tamanho um: as hipóteses individuais H_i .

Para ilustrar o PTF podemos analisar o diagrama a seguir, em que consideramos as hipóteses H_1, H_2, H_3 e H_4 e tomamos H_I^* como uma hipótese significativa.

$$\begin{array}{c}
 H_{1234}^* \\
 H_{123}^*, H_{124}^*, H_{134}^*, H_{234}^* \\
 H_{12}^*, H_{13}^*, H_{14}^*, H_{23}^*, H_{24}^*, H_{34}^* \\
 H_1^*, H_2^*, H_3, H_4
 \end{array}$$

Diante desta situação, observamos que apenas H_1 é rejeitada pelo PTF. Embora a hipótese H_2 seja significativa, não é rejeitada pelo PTF, pois nem todas intersecções H_J , com $\{2\} \subset J \subset \{1, 2, 3, 4\}$ são significantes.

Como visto em Wright (1992) o procedimento de teste fechado pode ser modificado de modo a gerar um valor-p ajustado para cada H_I . Isto é feito tomando como p_I o valor-p não ajustado da estatística T_I para testar H_I . A hipótese H_I é rejeitada somente se $p_J \leq \alpha$, para todo H_J , tal que $J \supseteq I$. Portanto, o valor-p ajustado para H_J deve ser o maior dos valores p_J . No caso geral, a fim de obter um valor-p ajustado para cada hipótese individual H_i seria necessário conduzir o teste e obter o valor-p não ajustado para cada possível subconjunto de hipóteses. O número total de testes que devem ser realizados é portanto $\sum_{i=1}^n \binom{n}{i} = 2^n - 1$. Felizmente, em certos casos existem atalhos para isto, tal como o procedimento de Holm que veremos adiante.

2.3 Procedimentos de Bonferroni

2.3.1 Bonferroni Clássico

A desigualdade de Bonferroni é frequentemente utilizada na realização de testes múltiplos a fim de definir um limite superior sobre o nível de significância conjunto α (Miller, 1981). O uso desta desigualdade na teoria dos testes múltiplos é usualmente chamada de procedimento de Bonferroni.

Seja $T_1, \dots, T_n$ um conjunto de n estatísticas para testar as hipóteses $H_1, \dots, H_n$ e $p_1, \dots, p_n$ os valores-p correspondentes a estas estatísticas. Pelo procedimento clássico

de Bonferroni, a hipótese nula $H_0 = \bigcap_{i=1}^n H_i$ é rejeitada se qualquer valor-p é menor que α^* . Além disso, a hipótese H_i específica é rejeitada para cada $p_i \leq \alpha^*$ com $i = 1, \dots, n$. Cada α^* é escolhido de modo que sua soma seja igual a α . Geralmente, os α^* preferidos são todos iguais a α/n , mas alocações diferentes podem ser utilizadas.

O método de Bonferroni controla a FWER. A verificação deste argumento está na sequência e segue quase que imediatamente da desigualdade de Bonferroni. Consideramos p_i o valor-p para testar H_i , tal que cada hipótese H_i pode ser vista como um subconjunto w_i de Ω . Assumimos especificamente que $P\{p \leq u\} \leq u$ para qualquer $u \in (0, 1)$ e qualquer $P \in w_i$. Notemos que não é necessário que a distribuição de p_i seja uniforme em $(0, 1)$ sempre que H_i é verdadeira.

Teorema 2.2 (Lehmann & Romano (2005)). *Se para cada $i = 1, \dots, n$, as hipóteses H_i são rejeitadas quando $p_i \leq \alpha/n$, então a FWER para os testes múltiplos de $H_1, \dots, H_n$ se mantém menor ou igual a α .*

Demonstração. Suponhamos que as hipóteses H_i com $i \in I$ sejam verdadeiras, com $|I|$ denotando a cardinalidade de I . Da desigualdade de Bonferroni temos que,

$$\begin{aligned} \text{FWER} &= P\{\text{rejeitar qualquer } H_i \text{ com } i \in I\} \leq \sum_{i \in I} P\{\text{rejeitar } H_i\} \\ &= \sum_{i \in I} P\left\{p_i \leq \frac{\alpha}{n}\right\} \leq \sum_{i \in I} \frac{\alpha}{n} \leq |I| \alpha/n \leq \alpha. \end{aligned}$$

□

Como vemos, este procedimento clássico controla satisfatoriamente a FWER. Todavia, em algumas situações é bastante conservativo, isto é, o valor da FWER real é às vezes muito menor que o nível α estabelecido.

O método de Bonferroni é um exemplo de teste de etapa única que tem como característica marcante o fato de poder ser aplicado em qualquer situação de inferências simultâneas. Contudo, nos métodos de uma etapa o controle da FWER tem a consequência de que quanto maior o número de hipóteses na família, menor é o poder para testar as hipóteses individuais. Procedimentos de etapas múltiplas superam esta desvantagem, e é por isso que os MCMs vistos como melhoramentos do método clássico de Bonferroni apresentam mais de uma etapa.

2.3.2 Procedimento de Bonferroni de rejeição sequencial

Holm (1979) propôs um método descendente que ficou conhecido na literatura como procedimento de Bonferroni de rejeição sequencial ou procedimento de Holm. Este método é definido por meio dos valores-p, $p_1, \dots, p_n$, dos n testes individuais.

Consideramos estes níveis descritivos dos testes ordenados denotados por $p_{(1)}, \dots, p_{(n)}$ e $H_{(1)}, \dots, H_{(n)}$ as hipóteses associadas. Dessa forma, o procedimento de Holm pode ser definido através das seguintes etapas:

Passo 1. Se $p_1 > \alpha/n$, as hipóteses $H_1, \dots, H_n$ não são rejeitadas e o procedimento termina. Se $p_1 \leq \alpha/n$, rejeitamos $H_{(1)}$ e o teste continua com $n - 1$ hipóteses testadas ao nível $\alpha/(n - 1)$;

Passo 2. Se $p_1 \leq \alpha/n$, mas $p_2 > \alpha/(n - 1)$, não rejeitamos $H_2, \dots, H_n$ e paramos. Se $p_1 \leq \alpha/n$ e $p_2 \leq \alpha/(n - 1)$, rejeitamos $H_{(2)}$ além de $H_{(1)}$ e o teste continua com $n - 2$ hipóteses ao nível $\alpha/(n - 2)$;

Em suma, o procedimento continua desta maneira rejeitando H_i quando

$$p_{(j)} \leq \alpha/(n - j + 1),$$

para todo $j = 1, \dots, n$.

O teorema dado na sequência mostra que o método de Holm controla a FWER ao nível α em associações livres (“*free combinations*”) das hipóteses. Esta definição de associações livres tem o objetivo de salientar que todos os subconjuntos de hipóteses nulas poderiam aparecer como conjunto de hipóteses verdadeiras. Pode haver situações em que nem todos os subconjuntos são permitidos por algum motivo, por exemplo, em situações em que a verdade de duas hipóteses implica a verdade ou a falsidade de uma terceira hipótese. É importante ser observado que um nível de significância múltiplo α para algumas combinações restritas impõe menos condições no procedimento de teste em relação ao nível de significância múltiplo em associações livres, ou seja, um procedimento de teste com nível múltiplo α em combinações livres mantém o nível múltiplo α para qualquer tipo de combinações restritas.

Teorema 2.3 (Lehmann & Romano (2005)). *No procedimento de Holm a FWER é menor*

ou igual a α .

Demonstração. Suponha H_i com $i \in I$ o conjunto das hipóteses verdadeiras. Assim, $P \in w_i$ se, e somente se, $i \in I$. Seja j o menor índice satisfazendo

$$p_j = \min_{i \in I} p_i.$$

Note que $j \leq n - |I| + 1$. O procedimento de Holm comete uma rejeição falsa se

$$p_{(1)} \leq \alpha/n, p_{(2)} \leq \alpha/(n-1), \dots, p_{(j)} \leq \alpha/(n-j+1),$$

implicando que

$$\min_{i \in I} p_i = p_{(j)} \leq \alpha/(n-j+1) \leq \alpha/|I|.$$

Logo, pela desigualdade de Bonferroni, a probabilidade de uma rejeição falsa é dada por

$$P \left\{ \min_{i \in I} p_i \leq \alpha/|I| \right\} \leq \sum_{i \in I} P \{p_i \leq \alpha/|I|\} \leq \sum_{i \in I} \alpha/|I| \leq \alpha.$$

□

Em quase todas as situações o procedimento de Holm tem probabilidade estritamente maior de rejeitar uma hipótese falsa em relação ao procedimento clássico de Bonferroni. O ganho de poder obtido usando o procedimento de Holm ao invés de Bonferroni depende muito da hipótese alternativa. O ganho é pequeno se todas as hipóteses forem “quase verdadeiras”, mas pode ser considerável se as hipóteses forem “completamente falsas” (Holm, 1979).

2.4 Método de Simes

Entre os procedimentos vistos como melhoramentos do método clássico de Bonferroni, um dos que recebeu bastante atenção ao longo dos anos e que damos grande importância neste trabalho foi proposto por Simes (1986).

Consideremos um conjunto de n hipóteses nulas, H_i ($i = 1, \dots, n$), e suponhamos que para cada hipótese individual temos uma estatística do teste apropriada, tal que os

valores-p oriundos dos testes das n hipóteses nulas são $p_1, \dots, p_n$. O método formulado por Simes tem o propósito de testar a hipótese nula global $H_0 = \bigcap_{i=1}^n H_i$ a um nível de significância α pré-estabelecido, em termos dos valores-p das estatísticas do teste, sugerindo a rejeição de H_0 se $p_{(i)} \leq i\alpha/n$ para ao menos um i , em que $p_{(1)}, \dots, p_{(n)}$ são os valores-p ordenados.

Devido à região de rejeição do método de Simes conter a região de rejeição do método de Bonferroni, este procedimento modificado é claramente mais poderoso que o método clássico. Também foi observado via simulação em Simes (1986), que esta melhoria do poder é bastante significativa quando as estatísticas dos testes são altamente correlacionadas. Em relação à FWER, foi provado que o procedimento modificado mantém o controle a um nível α pré-estabelecido apenas quando as estatísticas do teste são independentes. Contudo, Simes (1986) conjecturou via simulação que este controle se mantém quando as estatísticas são dependentes com algumas distribuições específicas: o valor absoluto da normal multivariada e um certo tipo de distribuição gama multivariada.

Sarkar (1998) provou analiticamente que a conjectura de Simes é verdadeira para uma classe geral de distribuições multivariadas com dependência positiva, caracterizada pela condição multivariada totalmente positiva de ordem 2 (MTP₂), e para algumas misturas de MTP₂. A definição e algumas propriedades do conceito MTP₂ estão na sequência e podem ser vistos detalhados em Karlin & Rinott (1980).

Seja $\mathcal{X} = \prod_{i=1}^n \mathcal{X}_i$ um produto de espaços totalmente ordenados $\mathcal{X}_i$, $i = 1, \dots, n$, com a ordenação parcial definida da seguinte maneira: para qualquer $\mathbf{x}, \mathbf{y} \in \mathcal{X}$, escrevemos $\mathbf{x} \leq \mathbf{y}$ se $\mathbf{x} = (x_1, \dots, x_n)$ e $\mathbf{y} = (y_1, \dots, y_n)$ satisfazem $x_i \leq y_i$ em $\mathcal{X}_i$, para $i = 1, \dots, n$. Consideramos $\mathbf{x} \vee \mathbf{y} = (\max(x_1, y_1), \dots, \max(x_n, y_n))$ e $\mathbf{x} \wedge \mathbf{y} = (\min(x_1, y_1), \dots, \min(x_n, y_n))$.

Definição 2.1. Uma função $f : \mathcal{X} \rightarrow [0, \infty)$ é dita ser MTP₂ (TP₂ quando $n = 2$) se para todo $\mathbf{x}, \mathbf{y} \in \mathcal{X}$,

$$f(\mathbf{x} \vee \mathbf{y})f(\mathbf{x} \wedge \mathbf{y}) \geq f(\mathbf{x})f(\mathbf{y}).$$

Um vetor n -dimensional, $\mathbf{X} = (X_1, \dots, X_n)$, ou sua distribuição é chamado MTP₂ se sua densidade é MTP₂.

Resultado 2.1. Se $f_1(x)$ e $f_2(x)$ são ambas MTP₂ em $\mathcal{X}$, então $f_1(x)f_2(x)$ é MTP₂.

Resultado 2.2. Se $f(x)$ é MTP_2 em $\mathcal{X}$, então

$$\int \dots \int f(x) \prod_{i=m+1}^n dx_i$$

é MTP_2 em $\prod_{i=1}^m \mathcal{X}_i$.

Agora, enunciamos de maneira análoga a Sarkar (1998) o teorema que leva à prova da conjectura de Simes.

Teorema 2.4. *Sejam $X_{(1)} \leq \dots \leq X_{(n)}$ as componentes ordenadas de um vetor aleatório MTP_2 , $\mathbf{X} = (X_1, \dots, X_n)$, e F_i a função de distribuição acumulada marginal de X_i . Então, temos as seguintes afirmações.*

(i) Para $a_1 \leq \dots \leq a_n$, fixados

$$P\{X_{(1)} \geq a_1, \dots, X_{(n)} \geq a_n\} \geq 1 - \frac{1}{n} \sum_{i=1}^n F_i(a_n), \quad (2.6)$$

se $F_i(a_j)/j$ é não decrescente em $j = 1, \dots, n$ para todo $i = 1, \dots, n$.

(ii) Para $b_1 \leq \dots \leq b_n$, fixados

$$P\{X_{(1)} \leq b_1, \dots, X_{(n)} \leq b_n\} \geq \frac{1}{n} \sum_{i=1}^n F_i(a_1), \quad (2.7)$$

se $\bar{F}_i(b_{n-j+1})/j$ é não decrescente em $j = 1, \dots, n$ para todo $i = 1, \dots, n$, em que $\bar{F}(x) = 1 - F(x)$.

Observação 2.1. Suponhamos agora que X_i s têm uma distribuição marginal comum F e que, para cada $i = 1, \dots, n$, o valor-p aleatório p_i correspondente a H_i é baseado em um teste unilateral à esquerda ou à direita em X_i . Desde que p_i é definido como $F(X_i)$ para um teste unilateral à esquerda e como $\bar{F}(X_i)$ para um teste unilateral à direita, em termos de p_i , as desigualdades (2.6) e (2.7) são equivalentes a

$$P\{p_{(i)} \geq c_i, i = 1, \dots, n\} \geq 1 - c_n,$$

em que $0 < c_1 \leq \dots \leq c_n < 1$ são tais que c_i/i é não decrescente em $i = 1, \dots, n$. Definindo $c_i = i\alpha/n$ para todo i , concluímos a prova da conjectura de Simes.

Karlin & Rinott (1980) deram uma lista de distribuições que satisfazem a condição (MTP₂) sob a hipótese nula. No entanto, vamos considerar uma subclasse dessas distribuições que satisfazem uma outra condição que pode ser vista em Sarkar & Chang (1997). Esta condição será a chave para mostrar que o método de Simes controla a FWER quando a estatística envolvida nos testes é a *log-rank*. Vamos descrever essas distribuições em termos das estatísticas dos testes, $X_1, \dots, X_n$.

Sejam $X_1, \dots, X_n$, variáveis aleatórias contínuas que, sob as hipóteses nulas são condicionalmente i.i.d. dada uma v.a. Z , com $X_i|Z = z \sim f(x_i, z)$, $i = 1, \dots, n$, em que $f(x_i, z)$ é TP₂ em (x_i, y) , e $g(z)$ é a densidade de Z . A densidade conjunta de $X_1, \dots, X_n$ sob a hipótese nula, que é da forma

$$\int \prod_{i=1}^n f(x_i, z) g(z) dz, \quad (2.8)$$

é MTP₂. Este fato segue dos Resultados (2.1 e 2.2).

Sarkar & Chang (1997) provaram que o método de Simes controla a FWER quando as estatísticas têm densidade conjunta da forma (2.8). Distribuições como estas aparecem muitas vezes em testes múltiplos. Por exemplo, a distribuição normal padrão multivariada equicorrelacionada, o valor absoluto da t multivariada central, a F multivariada central e o valor absoluto da normal padrão multivariada equicorrelacionada. Este último exemplo está detalhado adiante, mas para entendê-lo precisamos da próxima observação que pode ser vista em Curnow & Dunnett (1962).

Observação 2.2. Sejam $X_1, \dots, X_n$, n variáveis aleatórias normais padronizadas com coeficientes de correlação ρ_{ij} , e denotamos a função densidade de probabilidade (f.d.p) conjunta por $f(t_1, \dots, t_n)$. A função de distribuição acumulada (f.d.a) é definida como

$$\begin{aligned} F_n(x_i, \{\rho_{ij}\}) &= P\{X_i < x_i; \text{ para todo } i\} \\ &= \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \dots \int_{-\infty}^{x_n} f(t_1, t_2, \dots, t_n) dt_1 dt_2 \dots dt_n. \end{aligned} \quad (2.9)$$

Em algumas situações a integral (2.9) pode ser reduzida a uma integral simples, por exemplo, quando a matriz de correlação tem a estrutura $\rho_{ij} = \rho_i \rho_j$, para $i \neq j$ e $-1 \leq \rho_i \leq 1$. Dessa maneira, as variáveis X_i 's podem ser representadas em termos de $n + 1$

variáveis normais padronizadas independentes $Y_1, \dots, Y_n, Z$, de modo que

$$X_i = (1 - \rho)^{1/2} Y_i + \rho^{1/2} Z \quad \text{para } i=1, 2, \dots, n. \quad (2.10)$$

Como mencionado em Dunnett & Sobel (1955), quando $\rho_{ij} = \rho > 0$, temos a estrutura de correlação anterior satisfeita, uma vez que esta ocorre quando $\rho_i = \sqrt{\rho}$. Dessa forma, como os X_i 's são mutualmente independentes e utilizando a representação dada em (2.10), temos que

$$\begin{aligned} F_n(x_i, \{\rho\}) &= \int_{-\infty}^{\infty} P \left\{ Y_i < \frac{(x_i - \rho^{1/2} z)}{(1 - \rho)^{1/2}} \right\} f(z) dz \\ &= \int_{-\infty}^{\infty} \left\{ \prod_{i=1}^n \Phi \left(\frac{x_i - \rho^{1/2} z}{(1 - \rho)^{1/2}} \right) \right\} \phi(z) dz, \end{aligned} \quad (2.11)$$

em que $\Phi(\cdot)$ e $\phi(\cdot)$ representam respectivamente a f.d.a e a f.d.p da $N(0, 1)$. Assim, a densidade conjunta de $\{X_i\}$ é

$$\int_{-\infty}^{\infty} \prod_{i=1}^n \left\{ \frac{1}{(1 - \rho)^{1/2}} \phi \left(\frac{x_i - \rho^{1/2} z}{(1 - \rho)^{1/2}} \right) \right\} \phi(z) dz. \quad (2.12)$$

Exemplo 2.1 (Sarkar & Chang (1997)). A f.d.p da distribuição do valor absoluto da normal multivariada equicorrelacionada satisfaz (2.8), então a conjectura de Simes é válida quando as estatísticas dos testes têm esta distribuição.

Primeiramente suponha que temos $n + 1$ populações com distribuição normal independentes $N(\mu_i, \sigma^2)$, com média μ_i desconhecida e variância conhecida dada por σ^2 , com $i = 1, \dots, n$. Queremos testar as hipóteses

$$\begin{cases} H_0 : \cap_{i=1}^n H_i \\ \bar{H}_0 : \cup_{i=1}^n \bar{H}_i \end{cases}$$

em que, $H_i : \mu_i = \mu_0$ para todo i e $\bar{H}_i : \mu_i \neq \mu_0$ para ao menos um i , com base em amostras independentes, uma de tamanho m_0 de populações $N(\mu_0, \sigma^2)$ e amostras de tamanho m_1 para cada uma das outras n populações. Vamos considerar o valor absoluto da seguinte

estatística do teste para avaliar H_i contra $\bar{H}_i$

$$X_i = \left[\frac{(\bar{X}_i - \bar{X}_0)}{\sigma} \right] \left[\frac{m_0 m_1}{(m_0 + m_1)} \right]^{1/2}, \quad (2.13)$$

para $i = 1, \dots, n$, em que $\bar{X}_i$ é a i -ésima média amostral com $i = 0, \dots, n$. A distribuição de probabilidade conjunta desta estatística sob H_0 é o valor absoluto da normal multivariada equicorrelacionada com média zero, variância um e correlação $\rho = m_1/(m_1 + m_0)$. Embora ρ seja conhecido neste exemplo, nenhum valor específico é realmente necessário para mostrar que a densidade de probabilidade tem a forma (2.8). Além disso, como essa distribuição depende de ρ somente através do seu valor absoluto, podemos supor sem perda de generalidade que ρ é não negativo. Dessa maneira, usando a representação de $\{X_i\}$ como em (2.10), pode ser notado que condicionalmente, dado $Y_0 = z$, X_i^2 's são i.i.d como

$$(1 - \rho)\chi_1^2 \left(\frac{\rho z^2}{1 - \rho} \right),$$

em que $\chi_m^2(\lambda)$ denota uma variável aleatória qui-quadrada não central com m graus de liberdade e parâmetro de não centralidade λ , e $Y_0^2 \sim \chi_1^2$ representa uma variável aleatória qui-quadrada central com 1 grau de liberdade. Dessa forma, a densidade conjunta do valor absoluto da normal multivariada equicorrelacionada com média zero, variância um e correlação ρ é

$$\int_0^\infty \prod_{i=1}^n \left\{ \frac{2x_i}{1 - \rho} f \left(\frac{x_i^2}{1 - \rho}, \frac{\rho z}{1 - \rho} \right) \right\} g(z) dz \quad (2.14)$$

com $f(x, \lambda)$ representando a densidade de $\chi_1^2(\lambda)$ em x e $g(z)$ representando a densidade de χ_1^2 em z . Como $f(x, \lambda)$, com $x, \lambda \geq 0$, é conhecida por ser TP2 em (x, λ) (Karlin, 1968), concluímos que a densidade (2.14) é da forma (2.8). Com esse exemplo, vemos que quando as estatísticas do teste têm como distribuição o valor absoluto da normal multivariada equicorrelacionada, a conjectura de Simes é válida.

2.5 Método de Simes para testes individuais

Nesta seção enunciamos os métodos propostos por Hochberg (1988), Hommel (1988) e Rom (2013) que podem ser vistos como extensões do método de Simes para realizar inferências em hipóteses individuais uma vez que a hipótese global H_0 é rejeitada.

2.5.1 Método de Hochberg

O procedimento de etapas múltiplas proposto por Hochberg (1988) pode ser interpretado como uma versão do tipo ascendente do procedimento de Bonferroni. Enquanto o procedimento de Holm inicia o teste analisando o menor valor-p, o método de Hochberg inicia o teste observando o valor-p menos significativo. Este método é fundamentado no princípio dos testes fechados (Marcus *et al.*, 1976), o que intensifica o controle da FWER no sentido forte. Deve ser assumido neste procedimento que a coleção de hipóteses satisfaça a condição de combinações livres (Seção 2.3.2).

Sejam $p_{(1)}, \dots, p_{(n)}$, os valores-p ordenados de n estatísticas para testar as hipóteses $H_1, \dots, H_n$. Esta simples extensão do método de Simes pode ser ilustrada sucintamente através das seguintes etapas:

Passo 1. Se $p_{(n)} \leq \alpha$, então todas as hipóteses H_i com $i = 1, \dots, n$ são rejeitadas e paramos; caso contrário seguimos para o Passo 2.

Passo 2. Se $p_{(n-1)} \leq \alpha/2$, então todas as hipóteses $H_{(i)}$, com $i = 1, \dots, n-1$ são rejeitadas e paramos; caso contrário avançamos para o Passo 3.

$\vdots$

Passo n. Se $p_{(1)} \leq \alpha/n$, rejeitamos H_i , $i = 1$, e paramos.

Em diversas situações o método de Hochberg é mais poderoso em relação ao método de Holm. Contudo, por se tratar de uma extensão do método de Simes, para garantir o controle efetivo da FWER as estatísticas dos testes precisam ser independentes ou terem distribuição MTP_2 (ou certas misturas de MTP_2).

2.5.2 Método de Hommel

O procedimento proposto por Hommel (1988) testa as hipóteses individuais H_i por meio do teste de Simes como T_I no procedimento dos testes fechados. Os cálculos deste método são mais trabalhosos em relação ao método de Hochberg, porém o procedimento de testes fechados não precisa ser feito por completo. Para algum α fixado, Hommel (1989) mostrou que o seguinte caminho é suficiente. Seja j o número de hipóteses no

maior subconjunto de hipóteses para o qual o teste de Simes não é significativo. Ou seja,

$$j = \max\{i \in \{1, \dots, n\} : P_{(n-i+k)} > k\alpha/i \text{ para todo } k = 1, \dots, i\},$$

Se o máximo não existir, todos H_i ($i = 1, \dots, n$) são rejeitados; caso contrário, H_i é rejeitada quando $p_i \leq \alpha/j$. Este procedimento controla o nível múltiplo α desde que cada teste de Simes para H_I tem um nível de significância α . Por se tratar de uma extensão de Simes, este controle é mantido se as n estatísticas são independentes ou apresentam distribuição MTP₂ (ou certas misturas delas).

O cálculo dos valores-p ajustados para o procedimento de Hommel é mais simples de entender referindo-se diretamente ao procedimento de testes fechados, que quando realizado completamente exigiria que, para cada hipótese H_i , o valor-p do teste de Simes possa ser obtido para cada subconjunto de hipóteses contendo H_i . O valor-p ajustado por Hommel é então o maior desses valores-p do teste de Simes. No entanto, não é necessário testar cada subconjunto. Para subconjuntos contendo m das n hipóteses, é suficiente testar o único subconjunto contendo os maiores $(m - 1)$ valores-p além de p_i . Isto poderia ser chamado de “subconjunto menos significativo” de tamanho m que contém H_i . Claramente, se o teste de Simes é significativo ao nível α para esse subconjunto, será significativo ao nível α para todos os outros subconjuntos de tamanho m contendo H_i . O maior destes n valores-p de Simes é o valor-p ajustado de Hommel.

Na sequência ilustramos o algoritmo extraído de Wright (1992) que calcula o valor-p ajustado pelo método de Hommel. Seja $p_{(i)}$ os valores-p sem ajustes ordenados e $b_{(i)}$ o valor-p ajustado pelo método de Hommel.

Passo 1. Inicialmente defina $b_i = p_i$ para todo i ;

Passo 2. Para cada $m = n, (n - 1), \dots, 2$ (nesta ordem), faça o seguinte:

Passo 2a. Para $i > (n - m)$,

(i) Calcule o valor $c_i = (mp_i)/(m + i - n)$.

(ii) Encontre o menor dos valores c_i dados acima; e denote por $c_{\min}$.

(iii) Se $b_i < c_{\min}$, então temos $b_i = c_{\min}$.

Passo 2b. Para $i \leq (n - m)$,

- (i) Seja $c_i = \min(c_{\min}, mp_i)$.
- (ii) Se $b_i < c_i$, então temos $b_i = c_i$.

2.5.3 Método de Rom

O recente método proposto por Rom (2013) é uma modificação simples do método de Hochberg para realizar inferências sob as hipóteses individuais.

Consideramos o problema clássico de testar n hipóteses individuais, $H_1, \dots, H_n$, com $p_1, \dots, p_n$, os valores-p correspondentes. Assumimos que as hipóteses satisfazem a condição de combinações livres dada por Holm (1979) (Seção 2.3.2) e que os valores-p correspondentes sejam independentes. Sejam $p_{(1)}, \dots, p_{(n)}$ os valores-p ordenados. Este método pode ser descrito através das seguintes etapas:

Passo 1. Se $p_{(n)} \leq \alpha$, então todas as hipóteses são rejeitadas; caso contrário, H_n é retida e seguimos para o Passo 2.

Passo 2. Se $p_{(n-1)} \leq 3\alpha/2$, qualquer hipótese com valor-p menor ou igual a $\alpha/2$ é rejeitada; caso contrário, $H_{(n-1)}$ é retida e seguimos para o Passo 3.

Passo 3. Se $p_{(n-2)} \leq 4\alpha/6$, qualquer hipótese com valor-p menor ou igual a $\alpha/3$ é rejeitada, caso contrário $H_{(n-2)}$ é retida e seguimos para o próximo passo, de acordo com o nível crítico

$$C_i = (i+1)\alpha/(2i), \quad \text{com } i = 2, \dots, n.$$

Passo n. $p_{(1)}$ é comparado com α/n .

Rom (2013) mostrou através do princípio dos testes fechados que seu procedimento controla a FWER no sentido forte. Via simulação, foi verificado ainda que este procedimento tem uma ganho de poder maior em relação ao método de Hochberg e se comporta de maneira menos conservativa.

Capítulo 3

Modelo e Estatísticas dos Testes

Neste capítulo apresentamos o modelo discreto com censura aleatória que utilizamos neste trabalho e as estatísticas que foram empregadas nos testes: *log-rank* e Cramér-von Mises. Introduzimos ainda algumas notações e propriedades importantes.

3.1 Introdução

Análises estatísticas para dados discretos têm grande importância em diversas áreas tais como economia, medicina, educação e engenharia. Por exemplo, em economia do trabalho, quando queremos analisar a duração do desemprego, os dados disponíveis estão em meses completos ou em dias, e assim, a variável resposta é um valor inteiro. Na área de finanças, a duração de uma série de retornos positivos é usualmente analisada por períodos inteiros, como 5 ou 6 períodos. O mesmo ocorre na educação, o momento da evasão e o tempo de conclusão de um curso de graduação são medidos em semestres.

Quando observamos dados ao longo do tempo, uma situação frequente é a ocorrência de alguns dados somente em determinados períodos, isto é, sujeitos à censura. Para ilustrarmos isto, uma situação geral de censura à direita pode ser descrita se considerarmos $W_1^p, \dots, W_{n_p}^p$ variáveis aleatórias positivas independentes representando tempos de sobrevivência ou tempos de ocorrência de algum evento de interesse de n_p itens em uma população, com $p = 1, \dots, J$. Seja h^p a função intensidade da p -ésima população. Uma situação típica ocorre quando $\{W_m^p\}_{m=1}^{n_p}$ são censuradas à direita pelas variáveis

aleatórias positivas e independentes, $\{C_m^p\}_{m=1}^{n_p}$. Estas variáveis de censura, C_m^p , também são assumidas independentes de W_m^p . Dessa forma, nesse modelo geral de censura podemos observar que

$$X_m^p = \min\{W_m^p, C_m^p\}, \quad \delta_m^p = \mathbb{1}_{\{X_m^p = W_m^p\}},$$

com δ_m^p indicando se a variável W_m^p é censurada ou não.

Em um grande número de aplicações é de extrema importância testar a homogeneidade de populações na presença de censura. O problema está no desenvolvimento de métodos não paramétricos para testar a hipótese nula

$$H_0 = h^1(\ell) = \dots = h^J(\ell); \quad \ell \in \mathcal{X},$$

em que $\mathcal{X}$ é o domínio que codifica, por exemplo, o tempo de vida em um problema de análise de sobrevivência. O teste precisa ser consistente para uma grande classe de hipóteses alternativas.

Diante disto, realizamos neste trabalho os testes com as estatísticas *log-rank* e Cramér-von Mises. Esta escolha está fundamentada no fato de que a estatística *log-rank* é um dos métodos não paramétricos mais utilizados para testar H_0 sob censura aleatória. Contudo, esta estatística apresenta a desvantagem de não exibir um poder ao lidar com funções intensidades não proporcionais. Dessa forma, executamos também testes com a estatística Cramér-von Mises, que como pode ser visto em Leão & Ohashi (2011), é considerada ideal para modelos puramente discretos sob censura.

3.2 Modelo para os dados

Introduzimos nesta seção a notação básica e alguns fundamentos de inferência para o modelo discreto com censura utilizado neste trabalho. Maiores detalhes deste assunto podem ser vistos em Fleming & Harrington (1991) e em Leão & Ohashi (2011).

Consideramos $(\Omega, \mathcal{F}, \mathbb{P})$ um espaço de probabilidade e $\mathbb{N}$ o conjunto dos números naturais. Seja Y uma variável aleatória com valores em $\mathbb{N}$ e $\pi_i = \mathbb{P}[Y = i]$ denotando a probabilidade de falha da variável resposta (Y) na i -ésima categoria. Assumindo que as categorias são mutuamente exclusivas, temos que $\sum_{i=0}^{\infty} \pi_i = 1$. Assim, vemos que Y é

uma variável aleatória discreta.

Sejam W e C duas variáveis aleatórias discretas independentes com valores em $\mathbb{N}$, descrevendo respectivamente os eventos de interesse e de censura. Fixamos um número natural $J \geq 1$ e escrevemos $\mathcal{J} = \{1, \dots, J\}$. Denotamos $n^* := (n_1, \dots, n_J) \in \mathbb{N}^J$ e $N_J := \max\{n_1, \dots, n_J\}$. Definimos como n_q o tamanho da amostra aleatória correspondente à população, com $q \in \mathcal{J}$. Para um dado $p \in \mathcal{J}$, seja (W^p, C^p) uma população, tal que W^p e C^p são variáveis aleatórias discretas independentes que podem ser interpretadas como W e C respectivamente e seja X^p uma variável aleatória discreta definida por

$$X^p := W^p \wedge C^p = \min\{W^p, C^p\}.$$

Para qualquer $p \in \mathcal{J}$, tomamos as amostras aleatórias $\{(W_1^p, C_1^p), \dots, (W_{n_p}^p, C_{n_p}^p)\}$ da população (W^p, C^p) , tal que $1 \leq n_p < \infty$. Assumimos ainda que $\mathbb{P}[C^p = 0] = \mathbb{P}[W^p = 0] = 0$ para qualquer $p \in \mathcal{J}$. Assim, podemos introduzir

$$X_m^p := \min\{W_m^p, C_m^p\}, \quad V_m^p(i) := \mathbb{1}_{\{X_m^p \geq i\}}$$

e o processos de contagem

$$R_m^p(i) := \mathbb{1}_{\{X_m^p \leq i, X_m^p = W_m^p\}} \quad \text{e} \quad R_m^{C,p}(i) := \mathbb{1}_{\{X_m^p \leq i, X_m^p = C_m^p\}},$$

para $m = 1, \dots, n_p$ e $i \geq 0$. Os processos de contagem associados à p -ésima amostra aleatória são dados por

$$R^{n_p}(i) := \sum_{m=1}^{n_p} R_m^p(i) \quad \text{e} \quad R^{C,n_p}(i) := \sum_{m=1}^{n_p} R_m^{C,p}(i), \quad \text{e} \quad V^{n_p}(i) := \sum_{m=1}^{n_p} V_m^p(i); \quad i \geq 0.$$

Nesta configuração discreta, pode ser visto em Leão & Ohashi (2011) que a função intensidade h^p e o estimador de Kaplan-Meier $\hat{h}^{n_p}$ são dados respectivamente por

$$h^p(j) := \frac{\mathbb{P}[W^p = j]}{\mathbb{P}[W^p \geq j]}, \quad j \geq 0, \quad \text{e} \quad \hat{h}^{n_p}(i) = \frac{\Delta R^{n_p}(i)}{V^{n_p}(i)} \mathbb{1}_{\{V^{n_p}(i) > 0\}}; \quad i \geq 1.$$

3.3 Estatísticas dos Testes

3.3.1 Estatística *log-rank*

Nesta seção vamos estender a estatística log-rank ponderada proposta por Fleming & Harrington (1991) para comparar J populações discretas na presença de censura arbitrária.

Primeiramente vamos introduzir alguns objetos e definições básicas. Seja

$$V^{n^*}(i) := \sum_{k=1}^J V^{n_k}(i); \quad i \geq 0,$$

o número total de risco na categoria i , e para um certo $(q, q_1) \in \mathcal{J}^2$ e $\ell \geq 1$, denotamos

$$U_{n_{q_1}}^{n_q}(n^*, \ell) := \left(\frac{1}{n}\right)^{1/2} u(n^*, \ell) \left(\frac{V^{n_q}(\ell)V^{n_{q_1}}(\ell)}{V^{n^*}(\ell)}\right), \quad (3.1)$$

em que assumimos que o processo ponderado $u(n^*; \cdot)$ é limitado, converge em probabilidade para um valor real de uma função limitada e satisfaz certos pressupostos de mensurabilidade (Leão & Ohashi, 2011). Exemplos concretos de pesos são os processos ponderados de Harrington & Fleming (1982) e Tarone & Ware (1977) dados respectivamente por

$$\varphi\left(\frac{V^{n^*}(\ell)}{n}\right) \quad \text{e} \quad \left(\frac{\Delta R(\ell-1)}{V^{n^*}(\ell-1)}\right)^\beta \left(\prod_{j=0}^{\ell-1} \left(1 - \frac{\Delta R^{n^*}(j)}{V^{n^*}(j)}\right)\right)^\delta, \quad n^* \in \mathbb{N}^J, \ell \geq 1,$$

de modo que φ é uma função contínua limitada e (β, δ) são constantes positivas. Denotaremos a classe de processos ponderados da forma (3.1) por $\mathcal{K}$.

Na sequência, para um dado $q \in \mathcal{J}$, definimos $A_q := \{(x, y) \in \mathcal{J} \times \mathcal{J}; x \neq y, x \neq q, q \neq q, 1 \leq x < y \leq J\}$ e para $k \neq r$, definimos $A(k, r) := \{q_1 \in \mathcal{J}; q_1 \neq k, q_1 \neq r\}$. Para $n^* \in \mathbb{N}^J$ e $\ell \geq 1$, seja $\hat{Q}(n^*, \ell)$ uma matriz quadrada simétrica $\mathcal{J} \times \mathcal{J}$ dada por

$$[\hat{Q}(n^*, \ell)]_{rk} := \begin{cases} \hat{\phi}_{k, n^*}^2(\ell), & \text{se } k = r \\ \hat{\psi}_{n^*}(k, r, \ell), & \text{se } k \neq r, \end{cases}$$

em que

$$\begin{aligned}\hat{\phi}_{q,n^*}^2(\ell) &:= \sum_{q_1 \neq q} \left[\frac{|U_{n_{q_1}}^{n_q}(n^*, \ell)|^2}{V^{n_{q_1}}(\ell)} \hat{h}^{n_{q_1}}(\ell)[1 - \hat{h}^{n_{q_1}}(\ell)] + \frac{|U_{n_{q_1}}^{n_q}(n^*, \ell)|^2}{V^{n_q}(\ell)} \hat{h}^{n_q}(\ell)[1 - \hat{h}^{n_q}(\ell)] \right] \\ &+ 2 \sum_{(q_1, q_2) \in A_q} \frac{U_{n_{q_1}}^{n_q}(n^*, \ell) U_{n_{q_2}}^{n_q}(n^*, \ell)}{V^{n_q}(\ell)} \hat{h}^{n_q}(\ell)[1 - \hat{h}^{n_q}(\ell)], q \in \mathcal{J},\end{aligned}$$

e

$$\begin{aligned}\hat{\psi}_{n^*}(k, r, \ell) &:= \sum_{q_1 \in A(k, r)} \frac{U_{n_{q_1}}^{n_k}(n^*, \ell) U_{n_{q_1}}^{n_r}(n^*, \ell)}{V^{n_{q_1}}(\ell)} \hat{h}^{q_1}(\ell)[1 - \hat{h}^{q_1}(\ell)] \\ &- \sum_{q_1 \neq k} \frac{U_{n_{q_1}}^{n_k}(n^*, \ell) U_{n_k}^{n_r}(n^*, \ell)}{V^{n_k}(\ell)} \hat{h}^k(\ell)[1 - \hat{h}^k(\ell)] \\ &- \sum_{q_2 \neq r} \frac{U_{n_{q_2}}^{n_r}(n^*, \ell) U_{n_r}^{n_k}(n^*, \ell)}{V^{n_r}(\ell)} \hat{h}^r(\ell)[1 - \hat{h}^r(\ell)],\end{aligned}$$

se $r \neq k$ em $\mathcal{J}$.

Consideramos agora os seguintes limitantes das categorias

$$d^u = \sup \left\{ \ell : \min_{q \in \mathcal{J}} \theta^q(\ell) > 0 \right\}, \quad (3.2)$$

$$d_{n^*}^u = \sup \left\{ \ell : \min_{q \in \mathcal{J}} V^{n_q}(\ell) > 0 \right\}, n^* \in \mathbb{N}^J. \quad (3.3)$$

Podemos verificar que $d_{n^*}^u \rightarrow d^u$ em probabilidade com $n^* \rightarrow \infty$, tal que $1 \leq d^u \leq \infty$ e $d_{n^*}^u \leq \infty$ para cada $n^* \in \mathbb{N}^J$. Se $q \in \mathcal{J}$, então inspirados pela estatística *log-rank* ponderada proposta por Fleming & Harrington (1991) podemos introduzir a seguinte estatística linear geral para J -amostras.

$$\begin{aligned}LR_q(n^*, j) &:= \sum_{\ell=1}^j \sum_{q_1 \neq q} U_{n_{q_1}}^{n_q}(n^*, \ell) [\hat{h}^{n_q}(\ell) - \hat{h}^{n_{q_1}}(\ell)] \\ &= \sum_{\ell=1}^j \left(\frac{1}{n} \right)^{1/2} u(n^*, \ell) V^{n_q}(\ell) \left[\frac{\Delta R^{n_q}(\ell)}{V^{n_q}(\ell)} - \frac{\Delta R^{n^*}(\ell)}{V^{n^*}(\ell)} \right].\end{aligned}$$

Proposição 3.1. *Assumimos que U pertence à classe $\mathcal{K}$. Sob H_0 , o vetor aleatório*

$$LR(n^*, d_{n^*}^u) := (LR_1(n^*, d_{n^*}^u), \dots, LR_J(n^*, d_{n^*}^u))^T \quad (3.4)$$

converge em distribuição para uma variável Gaussiana multivariada $N(0, \Gamma(d^u))$ com $n^* \rightarrow \infty$ tal que a matriz de covariância $\Gamma(d^u)$ pode ser estimada com consistência por $\hat{\Gamma}(n^*, d_{n^*}^u)$, em que

$$\hat{\Gamma}(n^*, i) := \sum_{\ell=1}^i \hat{Q}(n^*, \ell), i \geq 1.$$

Similarmente ao caso contínuo, a soma dos componentes do vetor aleatório LR em (3.4) é nula. Assim, consideramos o vetor LR sem o último componente como na sequência

$$LR_0(n^*, d_{n^*}^u) := (LR_1(n^*, d_{n^*}^u), \dots, LR_{J-1}(n^*, d_{n^*}^u))^T \quad (3.5)$$

Dessa forma, definimos a estatística *log-rank* ponderada associada a LR_0 da seguinte maneira

$$X^2(n^*, d_{n^*}^u) := LR_0(n^*, d_{n^*}^u)^T \hat{\Gamma}_0(n^*, d_{n^*}^u)^{-1} LR_0(n^*, d_{n^*}^u), n^* \in \mathbb{N}^J, \quad (3.6)$$

em que $\hat{\Gamma}_0(n^*, \cdot)$ é dado por $\hat{\Gamma}(n^*, \cdot)$ mas sem a última linha e coluna. Pode ser verificado que a estatística (3.6) tem distribuição assintótica qui-quadrada com $J - 1$ graus de liberdade, em que $\hat{\Gamma}(n^*, d_{n^*}^u)^{-1}$ é a inversa usual.

A função poder pode ser obtida através da seguinte expressão

$$\begin{aligned} LR^q(i) &= \sum_{\ell=1}^i \sum_{q_1 \neq q} U_{n_{q_1}}^{n_q}(n^*, \ell) \left[\hat{h}^{n_q}(\ell) - \hat{h}^{n_{q_1}}(\ell) \right] = \\ &= \sum_{\ell=1}^i \sum_{q_1 \neq q} U_{n_{q_1}}^{n_q} \left[\frac{\Delta Y^{n_q}(\ell) \mathbb{1}_{\{V^{n_q}(\ell) > 0\}}}{V^{n_q}(\ell)} - \frac{\Delta Y^{n_{q_1}}(\ell) \mathbb{1}_{\{V^{n_{q_1}}(\ell) > 0\}}}{V^{n_{q_1}}(\ell)} \right] + \\ &= \sum_{\ell=1}^i \sum_{q_1 \neq q} U_{n_{q_1}}^{n_q} [h^q(\ell) - h^{q_1}(\ell)]. \end{aligned}$$

3.3.2 Condição MTP_2 para a estatística *log-Rank*

Nesta seção verificamos que a estatística *log-rank* ponderada satisfaz a condição MTP_2 . Alguns conceitos e resultados empregados na demonstração podem ser encontrados no Apêndice A.

Neste trabalho focamos no problema de MCMs de testar $J - 1$ populações com uma população controle ou referência. Consideramos a hipótese nula

$$H_q : h^q(\ell) = h^J(\ell), \forall \ell$$

e para qualquer $q = 1, \dots, J - 1$. Para cada hipótese individual construímos a estatística *log-rank* ponderada correspondente em valor absoluto,

$$X^{n_q}(i) = \frac{\left| \sum_{\ell=1}^i U_{n_J}^{n_q}(n^*, \ell) [\hat{h}^{n_q}(\ell) - \hat{h}^{n_J}(\ell)] \right|}{\sqrt{\sum_{\ell=1}^i \hat{\phi}_q^2(\ell)}}, \quad q = 1, \dots, J - 1,$$

com

$$\hat{\phi}_q^2(\ell) = \frac{|U_{n_J}^{n_q}(n^*, \ell)|^2}{V^{n_J}(\ell)} \hat{h}^{n_J}(\ell) [1 - \hat{h}^{n_J}(\ell)] + \frac{|U_{n_q}^{n_J}(n^*, \ell)|^2}{V^{n_q}(\ell)} \hat{h}^{n_q}(\ell) [1 - \hat{h}^{n_q}(\ell)]$$

e $U_{n_J}^{n_q}$ representando os processos ponderados. Da Observação A.2 segue que $X^{n_q}(i)$ tem como distribuição assintótica o valor absoluto de uma normal padrão.

Suponhamos que as $J - 1$ hipóteses nulas, $H_1, \dots, H_{J-1}$, tenham valores-p observáveis assintóticos, $p_1, \dots, p_{J-1}$, baseados em $X^{n_1}(i), \dots, X^{n_{J-1}}(i)$.

Teorema 3.1. *Sob a hipótese nula H_0 , a distribuição conjunta limitada de $\{X^{n_q}(i) : q = 1, \dots, J - 1\}$ satisfaz a condição MTP_2 .*

Demonstração. De acordo com o Exemplo 2.1, é suficiente mostrar que o vetor aleatório $\{X^{n_q}(i) : q = 1, \dots, J - 1\}$ tem como distribuição assintótica o valor absoluto de uma normal multivariada equicorrelacionada. Para isto, aplicamos alguns argumentos que podem ser vistos no Apêndice A (Lema A.2 e o Teorema A.2). Consideramos agora o *martingale array difference*

$$\xi_{m,q}^{n^*}(\ell) = U_{n_J}^{n_q}(n^*, \ell) \left[\frac{\Delta Y_m^q(\ell)}{V^{n_q}(\ell)} - \frac{\Delta Y_m^J(\ell)}{V^{n_J}(\ell)} \right], \quad q = 1, \dots, J - 1,$$

e para qualquer $a \in \mathbb{R}^{J-1}$,

$$\xi_m^{n^*}(\ell) = a_1 \xi_{m,1}^{n^*}(\ell) + \dots + a_{J-1} \xi_{m,J-1}^{n^*}(\ell), \quad \ell \geq 0 \text{ and } m = 1, \dots, N_J.$$

Então, temos que

$$\begin{aligned} \sum_{m=1}^{N_J} E [|\xi_m^*(\ell)|^2 | \mathcal{G}_{m-1}^*] &= \sum_{k=1}^{J-1} \sum_{m=1}^{N_J} |a_k|^2 E [|\xi_{m,k}^*(\ell)|^2 | \mathcal{G}_{m-1}^*] + \\ 2 \sum_{1 \leq r < k \leq J-1} \sum_{m=1}^{N_J} a_r a_k E [\xi_{m,r}^*(\ell) \xi_{m,k}^*(\ell) | \mathcal{G}_{m-1}^*] &= T_1(n^*, \ell) + T_2(n^*, \ell), \quad \ell \geq 1. \end{aligned}$$

Sabemos ainda que

$$T_1(n^*, \ell) \rightarrow \sum_{k=1}^{J-1} a_k \{ \alpha_{q,J}^J(\ell) h^q(\ell) [1 - h^q(\ell)] + \alpha_{q,J}^q(\ell) h^J(\ell) [1 - h^J(\ell)] \},$$

em probabilidade, para qualquer $\ell \geq 1$. Além disso, pode ser verificado que

$$\begin{aligned} T_2(n^*, \ell) &= 2 \sum_{1 \leq r < k \leq J-1} a_r a_k \frac{U_{n_J}^{n_r}(\ell) U_{n_J}^{n_k}(\ell)}{V^{n_J}(\ell)} \hat{h}^{n_J}(\ell) [1 - \hat{h}^{n_J}(\ell)] \rightarrow \\ 2 \sum_{1 \leq r < k \leq J-1} a_r a_k \gamma_J^{k,r}(\ell) h^{n_J}(\ell) [1 - h^{n_J}(\ell)], &\quad \text{em probabilidade.} \end{aligned}$$

Se considerarmos os processos ponderados de Tarone e Ware ou de Harrington e Fleming, podemos concluir de algumas expressões dadas no Apêndice (equações A.3, A.4 e A.8) que, sob H_0 , as constantes $\alpha_{q,J}^J(\ell) = \alpha_{q,J}^q(\ell) = \alpha(\ell)$ e $\gamma_J^{k,r}(\ell) = \gamma(\ell)$ são independentes dos índices k, r, J . Assim, sob $H_0 : h^1(\ell), \dots, h^J(\ell) = h(\ell)$, para todo $\ell = 1, \dots, i$, o vetor aleatório $\{\sum_{m=1}^{N_J} \xi_{m,q}^*(\ell) : q = 1, \dots, J-1\}$ tem distribuição assintótica com média zero e operador de covariância $\eta(i)$ dado por

$$\begin{aligned} \langle \eta(i), a \rangle_{J-1} &= \sum_{k=1}^{J-1} a_k \{ \alpha(\ell) h(\ell) [1 - h(\ell)] + \alpha(\ell) h(\ell) [1 - h(\ell)] \} + \\ 2 \sum_{1 \leq r < k \leq J-1} a_r a_k \gamma(\ell) h(\ell) [1 - h(\ell)] &= \\ \sum_{k=1}^{J-1} a_k \{ 2\alpha(\ell) h(\ell) [1 - h(\ell)] \} + 2 \sum_{1 \leq r < k \leq J-1} a_r a_k \gamma(\ell) h(\ell) [1 - h(\ell)]. \end{aligned}$$

Desde que $\hat{\phi}_q^2(\ell)$ é uma estimador consistente da variância, aplicamos os mesmos argumentos da prova do Teorema A.2 para obtermos que o vetor aleatório $\{X^{n_q}(i) : q = 1, \dots, J-1\}$ tem como distribuição assintótica o valor absoluto de uma normal multivariada com coeficiente de correlação

$$\rho = \frac{\sum_{\ell=1}^i \gamma(\ell) h(\ell) [1 - h(\ell)]}{\sum_{\ell=1}^i \alpha(\ell) h(\ell) [1 - h(\ell)]}$$

□

3.3.3 Estatística de Cramér-von Mises

Introduzimos agora a estatística de Cramér-von Mises para o modelo discreto geral induzido por X^p sob censura arbitrária à direita para infinitamente muitas categorias. Consideramos aqui as primeiras categorias observadas pela seguinte sequência de tempos de parada

$$d_{n^*}^\ell := \inf\{\ell : \Delta R^{n^*}(\ell) > 0\}, \quad (3.7)$$

tal que $R^{n^*}(i) := \sum_{k=1}^J R^{n_k}(i)$ é o número total de eventos de interesse para a categoria i . Podemos ver que $d_{n^*}^\ell \rightarrow d^\ell$ em probabilidade com $n^* \rightarrow \infty$ tal que

$$d^\ell := \inf\{\ell : b_1 h^1(\ell) + \dots + b_J h^J(\ell) > 0\}. \quad (3.8)$$

Diante disto, a estatística de Cramér-von Mises é definida por

$$CVM(n^*, d_{n^*}^\ell, d_{n^*}^u) := \sum_{k=d_{n^*}^\ell}^{d_{n^*}^u} \sum_{q=1}^{J-1} \hat{\phi}_{(q,n)}^2(k) L R_q^2(n^*, k), \quad n^* \in \mathbb{N}^J. \quad (3.9)$$

Em algumas situações é fundamental testar H_0 para categorias possivelmente infinitas, pois não sabemos a priori a existência de categorias maiores que i tal que, $\theta^q(i) = 0$ para cada $p \in \mathcal{J}$. Dessa maneira, é necessário assumir que $d^u \rightarrow \infty$ e portanto, somos forçados a trabalhar num espaço ℓ^2 de dimensão infinita constituído pelas sequências de quadrado somável de números reais. A estatística dada em (3.9) para possivelmente infinitas categorias ($d^u \leq \infty$) foi totalmente baseada na construção de um processo Gaussiano adequado, com valores no ℓ^2 e com operador de covariância $\mathcal{Y}_0(d^\ell, d^u) : \ell^2 \rightarrow \ell^2$ e autovalores $\{\lambda_{sq}; s \geq 1, q = 1, \dots, J-1\}$ que satisfazem

$$\sum_{s=1}^{\infty} \sum_{q=1}^{J-1} \lambda_{sq} < \infty.$$

A distribuição assintótica desta estatística não é livre de distribuição devido à presença inerente de saltos persistentes sobre $\mathbb{N}$. Na verdade, depende de $\{\lambda_{sq}; s \geq 1, q = 1, \dots, J-1\}$, que por sua vez depende dos dados. A fim de proporcionar um teste viável, construímos um estimador do operador $\mathcal{Y}_0(d^\ell, d^u)$ que pode ser obtido da maneira que está

na sequência. Introduzimos o conjunto aleatório de categorias observáveis $L(d_{n^*}^\ell, d_{n^*}^u) = \{d_{n^*}^\ell \leq \ell \leq d_{n^*}^u; \Delta \mathcal{R}^{n^*}(\ell) > 0\}$ e denotamos sua cardinalidade por $L(n^*)$. Para um dado $n^* \in \mathbb{N}^J$ e $a \in \ell^2$, definimos

$$\hat{\mathcal{Y}}_0(d_{n^*}^\ell, d_{n^*}^u)a$$

como uma sequência real, em que as $(J-1)L(n^*)$ -ésimas coordenadas são dadas por

$$\hat{\mathcal{Y}}_0(d_{n^*}^\ell, d_{n^*}^u)(a_1, \dots, a_{(J-1)L(n^*)}),$$

e $\hat{\mathcal{Y}}_0(d_{n^*}^\ell, d_{n^*}^u)$ é o operador autoadjunto definido pela forma quadrática sobre $\mathbb{R}^{(J-1)L(n^*)}$ a seguir

$$\begin{aligned} \langle \hat{\mathcal{Y}}_0(d_{n^*}^\ell, d_{n^*}^u)a, a \rangle &= \sum_{j \in L(d_{n^*}^\ell, d_{n^*}^u)} \langle \hat{M}_0(j) \hat{\Gamma}_0(j) \hat{M}_0(j) a_j, a_j \rangle_{\mathbb{R}^{J-1}} \\ &+ \sum_{\{\ell < j: \ell, j \in L(d_{n^*}^\ell, d_{n^*}^u)\}} \langle \hat{M}_0(j) \hat{\Gamma}_0(j) \hat{M}_0(j) a_j, a_j \rangle_{\mathbb{R}^{J-1}} \\ &+ \sum_{\{j < \ell: \ell, j \in L(d_{n^*}^\ell, d_{n^*}^u)\}} \langle \hat{M}_0(j) \hat{\Gamma}_0(j) \hat{M}_0(j) a_j, a_j \rangle_{\mathbb{R}^{J-1}}, \end{aligned}$$

em que $\hat{M}_0(\cdot) := \text{diag}(\hat{\phi}_{1,n^*}, \dots, \hat{\phi}_{J-1,n^*}(\cdot))$ e $a \in \mathbb{R}^{(J-1)L(n^*)}$. Portanto, $\hat{\mathcal{Y}}_0(d_{n^*}^\ell, d_{n^*}^u) : \ell^2 \rightarrow \ell^2$ é um sequência bem definida de operadores autoadjunto de postos finito.

Teorema 3.2. *Sejam $(d^\ell, d^u, d_{n^*}^\ell, d_{n^*}^u)$ as categorias e tempos de paradas definidos respectivamente por (3.8), (3.2), (3.7) e (3.3), e assumimos que o processo ponderado U pertença à classe $\mathcal{K}$. Então, sob H_0*

$$CVM(n^*, d_{n^*}^\ell, d_{n^*}^u) \rightarrow \sum_{s=1}^{\infty} \sum_{q=1}^{J-1} \lambda_{sq} \chi_{sq}^2 \quad \text{em distribuição com } n^* \rightarrow \infty,$$

tal que $\{\lambda_{sq}; s \geq 1, q = 1, \dots, J-1\}$ são os autovalores do operador de covariância $\mathcal{Y}_0(d^\ell, d^u)$. Se X^q é quadrado integrável, para cada $q \in \mathcal{J}$ temos que

$$\Lambda(n^*) := \sum_{s=1}^{L(n^*)} \sum_{q=1}^{J-1} \hat{\lambda}_{sq} \chi_{sq}^2 \mathbb{1}_{A(n^*)} \rightarrow \sum_{s=1}^{\infty} \sum_{q=1}^{J-1} \lambda_{sq} \chi_{sq}^2, \quad (3.10)$$

em distribuição com $n^* \rightarrow \infty$, em que $\{\hat{\lambda}_{sq}; s \leq s \leq L(n^*), q = 1, \dots, J-1\}$ são os

autovalores do estimador da covariância $\hat{\mathcal{Y}}_0(d_{n^*}^\ell, d_{n^*}^u)$ e

$$A(n^*) := \{\hat{\mathcal{Y}}_0(d_{n^*}^\ell, d_{n^*}^u) \text{ é não negativo}\},$$

de modo que $\mathbb{P}(A(n^*)) \rightarrow 1$ com $n^* \rightarrow \infty$.

Notamos de (3.10) que o valor p para testar a hipótese H_0 é dado por $\mathbb{P}[\Lambda(n^*) > CVM(n^*, d_{n^*}^\ell, d_{n^*}^u) | H_0]$. A lei de aproximação $\Lambda(n^*)$ é uma soma ponderada de variáveis aleatórias independentes e portanto, há algoritmos disponíveis para avaliar o valor-p que podem ser vistos em Duchesne & Lafaye De Micheaux (2010).

Capítulo 4

Estudo de simulação

Neste capítulo realizamos um estudo de simulação com o objetivo de comparar o comportamento dos procedimentos de Bonferroni, Holm, Hochberg, Hommel e Rom, em relação ao controle da FWER e ao ganho de poder quando as estatísticas envolvidas nos testes são a *log-rank* e a Cramér-von Mises. Focamos no problema de comparar uma população com uma população referência (tratamento versus controle). Trabalhamos com dados censurados provenientes de amostras aleatórias de distribuições discretas (Poisson) e contínuas (exponencial), mas enfatizamos o estudo com dados discretos, com os quais foram analisados três cenários distintos em relação à censura.

Toda a implementação foi desenvolvida em linguagem R (R Development Core Team, 2011) e para todas as circunstâncias tratadas neste capítulo adotamos o nível de significância múltiplo $\alpha = 0,05$.

Para averiguar o controle da FWER dos cinco MCMs escolhidos foram considerados nas simulações diferentes tamanhos de amostras (TA): 30, 50, 100, 150, 200, 250 e 300. Em todos os casos foram geradas dez populações da distribuição escolhida, sendo a primeira utilizada como referência (controle), resultando em nove hipóteses a serem testadas simultaneamente.

A fim de detectar minuciosamente as diferenças de comportamento entre os procedimentos de Bonferroni, Holm, Hochberg e Rom, avaliamos o poder de duas maneiras distintas. Na primeira estimativa do poder foram geradas 10 populações de distribuição Poisson (dados discretos) e exponencial (dados contínuos) da seguinte maneira, respecti-

vamente

$$X^{(1)}, \dots, X^{(9)} \sim P(\lambda) \quad \text{e} \quad X^{(10)} \sim P(\lambda + \delta)$$

e

$$X^{(1)}, \dots, X^{(9)} \sim E(\lambda) \quad \text{e} \quad X^{(10)} \sim E(\lambda + \delta),$$

em que δ iniciou com algum valor maior que zero e foi sendo incrementado até o poder atingir a taxa de 100%. A primeira população foi definida como referência, de modo que tivemos 9 hipóteses sendo testadas, sendo uma hipótese falsa.

A segunda estimativa do poder foi calculada por meio da razão entre o número de diferenças significativas detectadas pelo MCM e o número total de inferências. Consideramos os casos em que foram testadas 3, 5 e 10 hipóteses simultaneamente. Em cada situação criamos todas as configurações de hipóteses falsas, ou seja, quando testamos 3 hipóteses, analisamos os casos de 1, 2 e 3 hipóteses falsas presentes, e assim sucessivamente para os testes que envolveram 5 e 10 hipóteses.

Na primeira seção deste capítulo expomos as análises realizadas com dados discretos e na segunda seção está apresentado o estudo feito com dados contínuos.

4.1 Dados discretos

Para os estudos de simulações realizados com dados discretos foram geradas amostras aleatórias de distribuição Poisson e três cenários distintos em relação à censura foram analisados: no primeiro caso aproximadamente 40% dos dados foram censurados; no segundo caso 20% e por fim observamos o caso sem censura.

4.1.1 Controle da FWER

Uma vez que dados discretos empregados nas estatísticas *log-rank* e Cramér-von Mises não necessita de um tempo computacional muito grande, fixamos o número de iterações em 100.000 para podermos detectar da melhor maneira as diferenças de comportamento entre os MCMs em relação ao controle da FWER. Em cada caso geramos uma amostra aleatória de 10 populações com distribuição Poisson com parâmetros

$\lambda_1, \dots, \lambda_{10} = 10$, sendo a primeira população definida como controle (referência). Criamos as situações em que a censura foi fixada em 11 (40% dos dados censurados), 13 (20% dos dados censurados) e o caso sem censura.

Os valores e o comportamento da FWER em função dos diferentes tamanhos de amostras (TA) das cinco estratégias de comparações múltiplas empregadas nos testes das estatísticas *log-rank* e Cramér-von Mises dos três cenários de censura escolhidos estão ilustrados na sequência através das Tabelas 4.1, 4.2 e 4.3 e por meio das Figuras 4.1, 4.2 e 4.3.

TABELA 4.1: FWER dos 5 MCMs empregados nos testes das estatísticas *log-rank* e Cramér-von Mises para diferentes TA com censura fixada em 11.

Estatística <i>log-rank</i>							
Estratégia	TA=30	TA=50	TA=100	TA=150	TA=200	TA=250	TA=300
Bonferroni	0,04782	0,04823	0,04837	0,04799	0,04832	0,04799	0,04902
Holm	0,04782	0,04823	0,04837	0,04799	0,04832	0,04799	0,04902
Hochberg	0,04783	0,04823	0,04838	0,04801	0,04834	0,04801	0,04903
Hommel	0,04813	0,04846	0,04864	0,04822	0,04857	0,04823	0,04928
Rom	0,04906	0,04923	0,04958	0,04904	0,04926	0,04902	0,05001
Estatística Cramér-von Mises							
Estratégia	TA=30	TA=50	TA=100	TA=150	TA=200	TA=250	TA=300
Bonferroni	0,05093	0,04917	0,04941	0,04736	0,04985	0,04880	0,04951
Holm	0,05093	0,04917	0,04941	0,04736	0,04985	0,04880	0,04951
Hochberg	0,05096	0,04918	0,04943	0,04736	0,04985	0,04881	0,04952
Hommel	0,05122	0,04949	0,04967	0,04766	0,05015	0,04903	0,04981
Rom	0,05206	0,05044	0,05046	0,04853	0,05097	0,05078	0,05065

FIGURA 4.1: Gráficos do comportamento da FWER com censura fixada em 11.

TABELA 4.2: FWER dos 5 MCMs empregados nos testes das estatísticas *log-rank* e Cramér-von Mises para diferentes TA com censura fixada em 13.

Estatística <i>log-rank</i>							
Estratégia	TA=30	TA=50	TA=100	TA=150	TA=200	TA=250	TA=300
Bonferroni	0,05500	0,05080	0,05100	0,04993	0,04970	0,04902	0,04964
Holm	0,05500	0,05080	0,05100	0,04993	0,04970	0,04902	0,04964
Hochberg	0,05500	0,05082	0,05102	0,04994	0,04975	0,04902	0,04965
Hommel	0,05532	0,05117	0,05128	0,05113	0,05003	0,04921	0,04989
Rom	0,05643	0,05221	0,05222	0,05021	0,05100	0,05008	0,05002

Estatística Cramér von Mises							
Estratégia	TA=30	TA=50	TA=100	TA=150	TA=200	TA=250	TA=300
Bonferroni	0,05021	0,05099	0,04872	0,04984	0,04826	0,04884	0,04858
Holm	0,05021	0,05099	0,04872	0,04984	0,04286	0,04884	0,04858
Hochberg	0,05025	0,05103	0,04877	0,04984	0,04829	0,04885	0,04859
Hommel	0,05051	0,05127	0,04907	0,05004	0,04855	0,04910	0,04884
Rom	0,05159	0,05220	0,05013	0,05089	0,04936	0,04988	0,04955

FIGURA 4.2: Gráficos do comportamento da FWER com censura fixada em 13.

TABELA 4.3: FWER dos 5 MCMs empregados nos testes das estatísticas *log-rank* e Cramér-von Mises para diferentes TA com dados sem censura.

Estatística <i>log-rank</i>							
Estratégia	TA=30	TA=50	TA=100	TA=150	TA=200	TA=250	TA=300
Bonferroni	0,04516	0,04711	0,04793	0,04924	0,04902	0,04845	0,04840
Holm	0,04516	0,04711	0,04793	0,04924	0,04902	0,04845	0,04840
Hochberg	0,04517	0,04711	0,04794	0,04925	0,04903	0,04845	0,04842
Hommel	0,04541	0,04735	0,04814	0,04948	0,04918	0,04865	0,04864
Rom	0,04630	0,04818	0,04902	0,05013	0,04984	0,04942	0,04946

Estatística Cramér von Mises							
Estratégia	TA=30	TA=50	TA=100	TA=150	TA=200	TA=250	TA=300
Bonferroni	0,04744	0,04929	0,05006	0,04840	0,04855	0,04871	0,04851
Holm	0,04744	0,04929	0,05006	0,04840	0,04855	0,04871	0,04851
Hochberg	0,04746	0,04933	0,05007	0,04842	0,04856	0,04876	0,04855
Hommel	0,04778	0,04962	0,05034	0,04862	0,04887	0,04897	0,04878
Rom	0,04875	0,05042	0,05126	0,04937	0,04956	0,04984	0,04953

FIGURA 4.3: Gráficos do comportamento da FWER para dados sem censura.

Podemos observar por das meio das Figuras 4.1, 4.2 e 4.3, que em quase todos os cenários de censura trabalhados nos testes das duas estatísticas envolvidas, o procedimento de Rom apresentou os maiores valores para a FWER, que por um lado é bom, pois reafirma o fato argumentado em Rom (2013) de que este método é pouco conservativo; porém em certos casos de amostras pequenas ele apresentou uma taxa de erro elevada com respeito aos demais procedimentos.

No cenário de censura igual a 11, que equivale a 40% dos dados censurados (Figura 4.1), os métodos aplicados na estatística *log-rank* tiveram um comportamento

quase sem oscilações e atingiram o valor bem próximo a 5% com TA igual a 300. Já no emprego da estatística Cramér-von Mises, observamos maiores alterações, principalmente com amostras de tamanho 150, em que as 5 estratégias apresentaram taxas de erro menores em relação às demais.

Pela Figura 4.2, que ilustra o cenário da censura fixada em 13, podemos ver que para TA 30 e 50 os cinco procedimentos apresentaram taxas altas, mas foram atingindo o valor próximo a 5% conforme os TA foram aumentando.

Na situação analisada sem presença de censura (Figura 4.3) notamos que nos testes da estatística *log-rank* os 5 procedimentos apresentaram-se de maneira conservativa para TA pequenas. Já no teste da estatística de Cramér-von Mises os valores pequenos para FWER só ocorrem para TA 30, depois as taxas foram mantidas próximas a 5%, o nível nominal adotado.

4.1.2 Poder dos testes

Na primeira estimativa do poder dos cinco MCMs empregados, o número de iterações foi fixado em 10000 e em cada caso foram geradas amostras de tamanho 100 de 10 populações de distribuição Poisson com parâmetros

$$X^{(1)}, \dots, X^{(9)} \sim P(10) \quad \text{e} \quad X^{(10)} \sim P(10 + \delta),$$

tal que δ assumiu os valores 0, 3; 0, 5; 1; 1, 5; 2; 2, 5; 3; 3, 5; 4; 4, 5 e 5.

Averiguamos o poder considerando os três cenários de censura mencionados: censura igual a 11, 13 e o caso sem censura. Nas Tabelas 4.4, 4.5 e 4.6 estão os valores do poder em função do incremento de δ . As Figuras 4.4, 4.5 e 4.6 trazem os gráficos com a função poder dos cinco MCMs empregados nas estatísticas *log-rank* e Cramér-von Mises.

TABELA 4.4: Poder dos 5 MCMs empregados nas estatísticas *log-rank* e Cramér-von Mises em função do incremento de δ , censura fixada em 11.

Estatística <i>log-rank</i>					
δ	Bonferroni	Holm	Hochberg	Hommel	Rom
0,3	0,0136	0,0136	0,0136	0,0136	0,0574
0,5	0,0308	0,0308	0,0309	0,0311	0,0776
1	0,1557	0,1560	0,1561	0,1565	0,1980
1,5	0,4256	0,4264	0,4264	0,4267	0,4561
2	0,7015	0,7024	0,7026	0,7039	0,7196
2,5	0,9015	0,9017	0,9018	0,9021	0,9081
3	0,9704	0,9705	0,9705	0,9706	0,9723
3,5	0,9950	0,9950	0,9950	0,9950	0,9952
4	0,9990	0,9990	0,9990	0,9991	0,9991
4,5	0,9998	0,9998	0,9998	0,9998	0,9998
5	1,0000	1,0000	1,0000	1,0000	1,0000

Estatística Cramér-von Mises					
δ	Bonferroni	Holm	Hochberg	Hommel	Rom
0,3	0,0145	0,0146	0,0146	0,0147	0,0580
0,5	0,0376	0,0376	0,0376	0,0377	0,0825
1	0,1920	0,1924	0,1925	0,1930	0,2312
1,5	0,4944	0,4950	0,4950	0,4957	0,5213
2	0,7922	0,7930	0,7931	0,7939	0,8040
2,5	0,9015	0,9017	0,9018	0,9021	0,9081
3	0,9900	0,9901	0,9901	0,9901	0,9910
3,5	0,9982	0,9982	0,9982	0,9982	0,9984
4	0,9999	0,9999	0,9999	0,9999	0,9999
4,5	1,0000	1,0000	1,0000	1,0000	1,0000
5	1,0000	1,0000	1,0000	1,0000	1,0000

FIGURA 4.4: Gráficos da função poder dos 5 MCMs com censura fixada em 11.

TABELA 4.5: Poder dos 5 MCMs empregados nas estatísticas *log-rank* e Cramér-von Mises em função do incremento de δ , censura fixada em 13.

Estatística <i>log-rank</i>					
δ	Bonferroni	Holm	Hochberg	Hommel	Rom
0,3	0,0158	0,0159	0,0159	0,0159	0,0650
0,5	0,0436	0,0437	0,0438	0,0439	0,0851
1	0,2087	0,2094	0,2094	0,2098	0,2493
1,5	0,5409	0,5416	0,5416	0,5423	0,5672
2	0,8288	0,8291	0,8292	0,8297	0,8386
2,5	0,9675	0,9676	0,9676	0,9676	0,9697
3	0,9966	0,9966	0,9966	0,9967	0,9968
3,5	0,9997	0,9997	0,9997	0,9997	0,9997
4	1,0000	1,0000	1,0000	1,0000	1,0000

Estatística Cramér-von Mises					
δ	Bonferroni	Holm	Hochberg	Hommel	Rom
0,3	0,0170	0,0170	0,0170	0,0170	0,0617
0,5	0,0439	0,0441	0,0441	0,0443	0,0906
1	0,2254	0,2259	0,2260	0,2263	0,2655
1,5	0,5754	0,5764	0,5764	0,5767	0,5991
2	0,8668	0,8673	0,8673	0,8675	0,8743
2,5	0,9781	0,9782	0,9783	0,9785	0,9796
3	0,9985	0,9985	0,9985	0,9985	0,9985
3,5	0,9999	0,9999	0,9999	0,9999	0,9999
4	1,0000	1,0000	1,0000	1,0000	1,0000

FIGURA 4.5: Gráficos da função poder dos 5 MCMs com censura fixada em 13.

TABELA 4.6: Poder dos 5 MCMs empregados nas estatísticas *log-rank* e Cramér-von Mises em função do incremento de δ , dados sem censura.

Estatística <i>log-rank</i>					
δ	Bonferroni	Holm	Hochberg	Hommel	Rom
0,3	0,0159	0,0162	0,0162	0,0163	0,0635
0,5	0,0378	0,0379	0,0380	0,0382	0,0855
1	0,2410	0,2414	0,2415	0,2417	0,2807
1,5	0,6141	0,6144	0,6144	0,6149	0,6347
2	0,8951	0,8954	0,8955	0,8957	0,9016
2,5	0,9870	0,9870	0,9870	0,9870	0,9880
3	0,9990	0,9990	0,9990	0,9990	0,9992
3,5	1,0000	1,0000	1,0000	1,0000	1,0000
4	1,0000	1,0000	1,0000	1,0000	1,0000

Estatística Cramér-von Mises					
δ	Bonferroni	Holm	Hochberg	Hommel	Rom
0,3	0,0170	0,0170	0,0170	0,0170	0,0600
0,5	0,0430	0,0432	0,0432	0,0433	0,0873
1	0,2392	0,2397	0,2398	0,2403	0,2800
1,5	0,6290	0,6297	0,6299	0,6311	0,6491
2	0,9057	0,9060	0,9060	0,9065	0,9117
2,5	0,9889	0,9889	0,9889	0,9889	0,9897
3	0,9995	0,9996	0,9996	0,9996	0,9997
3,5	0,9999	0,9999	0,9999	0,9999	0,9999
4	1,0000	1,0000	1,0000	1,0000	1,0000

FIGURA 4.6: Gráficos da função poder dos 5 MCMs com dados sem censura.

Pelas Figuras 4.4, 4.5 e 4.6 é notório o fato de que os métodos de Bonferroni, Holm, Hochberg, Hommel e Rom apresentam praticamente o mesmo poder, exceto para valores bem pequenos de δ , em que o método de Rom apresenta superioridade em relação

aos demais.

Na situação de censura igual a 11 (Tabela 4.4), o poder das estratégias convergiu para exatamente 1 quando δ assumiu os valores 4, 5 e 5 nos testes das estatísticas *log-rank* e Cramér-von Mises, respectivamente. No cenário de censura igual a 13 (Tabela 4.5) esta convergência ocorreu com δ igual a 4 para as duas estatísticas. Já no caso sem censura (Tabela 4.6), o poder dos testes atingiu 100% com $\delta = 3, 5$ na estatística *log-rank* e $\delta = 4$ na estatística Cramér-von Mises.

A segunda avaliação do poder foi também realizada nos três cenários de censura escolhidos. Mantivemos o número de iterações em 10000 e tamanhos de amostras igual a 100. Exploramos as situações em que foram geradas 4, 6 e 11 populações de distribuição Poisson e sempre a primeira população foi tomada como referência. As expressões dadas na sequência ilustram a maneira como foram geradas as 4 populações para testarmos três hipóteses variando o número de hipóteses falsas:

$$\begin{aligned} X^{(1)}, X^{(2)}, X^{(3)} &\sim P(\lambda) \quad \text{e} \quad X^{(4)} \sim P(\lambda + \delta), \\ X^{(1)}, X^{(2)} &\sim P(\lambda) \quad \text{e} \quad X^{(3)}, X^{(4)} \sim P(\lambda + \delta), \\ X^{(1)} &\sim P(\lambda) \quad \text{e} \quad X^{(2)}, X^{(3)}, X^{(4)} \sim P(\lambda + \delta), \end{aligned}$$

em que $\lambda = 10$ e $\delta = 2$. De modo análogo foram geradas as demais populações para realizar 5 e 10 testes simultâneos.

As Tabelas 4.7, 4.8, 4.9, 4.10, 4.11 e 4.12 apresentam os valores do poder dos cinco MCMs empregados nas estatísticas *log-rank* e Cramér-von Mises para as diferentes configurações de censura adotadas. Nas Figuras 4.7, 4.8, 4.9, 4.10, 4.11 e 4.12 estão ilustrados o comportamento do poder das estratégias quando foram testadas 3, 5 e 10 hipóteses em função do número de hipóteses falsas.

TABELA 4.7: Poder dos 5 MCMs empregados na estatística *log-rank* em função do número de hipóteses falsas, censura fixada em 11.

Número de hipóteses	Número de hipóteses falsas	Estratégia				
		Bonferroni	Holm	Hochberg	Hommel	Rom
3	1	0,2889	0,2944	0,2956	0,2959	0,2962
	2	0,5552	0,5867	0,5893	0,5898	0,5900
	3	0,8277	0,9013	0,9067	0,9069	0,9070
5	1	0,1622	0,1641	0,1642	0,1646	0,1650
	2	0,3159	0,3271	0,3277	0,3286	0,3294
	3	0,4715	0,5014	0,5040	0,5061	0,5070
	4	0,6253	0,6865	0,6918	0,6944	0,6953
	5	0,7802	0,8922	0,9023	0,9029	0,9032
10	1	0,0746	0,0751	0,0751	0,0753	0,0756
	2	0,1436	0,1459	0,1459	0,1463	0,1471
	3	0,2119	0,2190	0,2190	0,2199	0,2213
	4	0,2819	0,2945	0,2948	0,2965	0,2989
	5	0,3514	0,3731	0,3739	0,3766	0,3796
	6	0,4226	0,4572	0,4585	0,4626	0,4668
	7	0,4881	0,5400	0,5433	0,5494	0,5544
	8	0,5589	0,6336	0,6387	0,6470	0,6517
	9	0,6272	0,7334	0,7427	0,7518	0,7558
	10	0,7001	0,8505	0,8738	0,8789	0,8808

FIGURA 4.7: Gráficos do poder em função do nº de hipóteses falsas dos 5 MCMs empregados na estatística *log-rank* com censura fixada em 11.

TABELA 4.8: Poder dos 5 MCMs empregados na estatística Cramér-von Mises em função do número de hipóteses falsas, censura fixada em 11.

Número de hipóteses	Número de hipóteses falsas	Estratégia				
		Bonferroni	Holm	Hochberg	Hommel	Rom
3	1	0,3057	0,3118	0,3125	0,3128	0,3129
	2	0,5958	0,6225	0,6253	0,6257	0,6259
	3	0,8873	0,9439	0,9467	0,9467	0,9468
5	1	0,1792	0,1812	0,1814	0,1816	0,1820
	2	0,3459	0,3559	0,3566	0,3574	0,3580
	3	0,5149	0,5409	0,5427	0,5440	0,5446
	4	0,6805	0,7354	0,7384	0,7394	0,7398
	5	0,8475	0,9371	0,9425	0,9427	0,9428
10	1	0,0836	0,0841	0,0841	0,0842	0,0845
	2	0,1602	0,1628	0,1628	0,1631	0,1638
	3	0,2382	0,2449	0,2450	0,2457	0,2467
	4	0,3171	0,3296	0,3298	0,3309	0,3324
	5	0,3937	0,4148	0,4151	0,4168	0,4189
	6	0,4728	0,5057	0,5068	0,5093	0,5120
	7	0,5502	0,5984	0,6001	0,6036	0,6062
	8	0,6288	0,6988	0,7025	0,7072	0,7096
	9	0,7077	0,8050	0,8118	0,8161	0,8178
	10	0,7840	0,9195	0,9330	0,9349	0,9355

FIGURA 4.8: Gráficos do poder em função do n° de hipóteses falsas dos 5 MCMs empregados na estatística Cramér-von Mises com censura fixada em 11.

TABELA 4.9: Poder dos 5 MCMs empregados na estatística *log-rank* em função do número de hipóteses falsas, censura fixada em 13.

Número de hipóteses	Número de hipóteses falsas	Estratégia				
		Bonferroni	Holm	Hochberg	Hommel	Rom
3	1	0,3168	0,3230	0,3237	0,3240	0,3241
	2	0,6181	0,6424	0,6436	0,6438	0,6438
	3	0,9172	0,9630	0,9650	0,9650	0,9650
5	1	0,1829	0,18478	0,18486	0,18518	0,1856
	2	0,35856	0,36756	0,36814	0,36878	0,36922
	3	0,5348	0,5583	0,5594	0,5601	0,5605
	4	0,7103	0,7557	0,7574	0,7581	0,7583
10	5	0,8787	0,9551	0,9585	0,9588	0,9588
	1	0,0878	0,0883	0,0884	0,0885	0,0887
	2	0,1696	0,1722	0,1723	0,1726	0,1730
	3	0,2531	0,2591	0,2592	0,2596	0,2605
	4	0,3340	0,3458	0,3460	0,3468	0,3481
	5	0,4160	0,4361	0,4366	0,4379	0,4394
	6	0,4980	0,5280	0,5287	0,5304	0,5325
	7	0,5823	0,6268	0,6279	0,6303	0,6320
	8	0,6618	0,7244	0,7271	0,7300	0,7317
	9	0,7453	0,8334	0,8382	0,8407	0,8419
	10	0,8253	0,9466	0,9548	0,9556	0,9558

FIGURA 4.9: Gráficos do poder em função do nº de hipóteses falsas dos 5 MCMs empregados na estatística *log-rank* com censura fixada em 13.

TABELA 4.10: Poder dos 5 MCMs empregados na estatística Cramér-von Mises em função do número de hipóteses falsas, censura fixada em 13.

Número de hipóteses	Número de hipóteses falsas	Estratégia				
		Bonferroni	Holm	Hochberg	Hommel	Rom
3	1	0,3242	0,3316	0,3321	0,3321	0,3322
	2	0,6276	0,6517	0,6525	0,6526	0,6528
	3	0,9375	0,9747	0,9759	0,9759	0,9759
5	1	0,1903	0,1925	0,1928	0,1929	0,1931
	2	0,3693	0,3780	0,3783	0,3787	0,3791
	3	0,5515	0,5721	0,5729	0,5734	0,5735
	4	0,7293	0,7700	0,7716	0,7720	0,7721
	5	0,9103	0,9710	0,9728	0,9729	0,9729
10	1	0,0907	0,0912	0,0913	0,0914	0,0915
	2	0,1770	0,1791	0,1792	0,1794	0,1798
	3	0,2628	0,2683	0,2684	0,2687	0,2694
	4	0,3484	0,3591	0,3592	0,3597	0,3606
	5	0,4346	0,4517	0,4519	0,4530	0,4539
	6	0,5191	0,5464	0,5471	0,5483	0,5495
	7	0,6049	0,6440	0,6453	0,6470	0,6486
	8	0,6925	0,7478	0,7496	0,7515	0,7526
	9	0,7766	0,8540	0,8573	0,8590	0,8597
	10	0,8650	0,9661	0,9703	0,9706	0,9707

FIGURA 4.10: Gráficos do poder em função do n° de hipóteses falsas dos 5 MCMs empregados na estatística Cramér-von Mises com censura fixada em 13.

TABELA 4.11: Poder dos 5 MCMs empregados na estatística *log-rank* em função do número de hipóteses falsas, dados sem censura.

Número de hipóteses	Número de hipóteses falsas	Estratégia				
		Bonferroni	Holm	Hochberg	Hommel	Rom
3	1	0,3292	0,3353	0,3356	0,3356	0,3357
	2	0,6401	0,6595	0,6604	0,6604	0,6604
	3	0,9499	0,9799	0,9806	0,9806	0,9806
5	1	0,1938	0,1959	0,1960	0,1961	0,1964
	2	0,3778	0,3858	0,3860	0,3866	0,3869
	3	0,5600	0,5790	0,5795	0,5799	0,5802
	4	0,7449	0,7802	0,7810	0,7813	0,7814
	5	0,9268	0,9789	0,9803	0,9803	0,9803
10	1	0,0931	0,0937	0,0937	0,0938	0,0940
	2	0,1823	0,1844	0,1845	0,1846	0,1850
	3	0,2685	0,2737	0,2737	0,2742	0,2748
	4	0,3588	0,3684	0,3685	0,3690	0,3698
	5	0,4463	0,4624	0,4626	0,4632	0,4642
	6	0,5337	0,5583	0,5587	0,5597	0,5608
	7	0,6208	0,6566	0,6574	0,6587	0,6596
	8	0,7118	0,7609	0,7622	0,7635	0,7642
	9	0,8011	0,8667	0,8696	0,8707	0,8709
	10	0,8886	0,9754	0,9787	0,9789	0,9789

FIGURA 4.11: Gráficos do poder em função do n° de hipóteses falsas dos 5 MCMs empregados na estatística *log-rank* com dados sem censura.

TABELA 4.12: Poder dos 5 MCMs empregados na estatística Cramér-von Mises em função do número de hipóteses falsas, dados sem censura.

Número de hipóteses	Número de hipóteses falsas	Estratégia				
		Bonferroni	Holm	Hochberg	Hommel	Rom
3	1	0,3304	0,3364	0,3370	0,3370	0,3370
	2	0,6427	0,6629	0,6634	0,6635	0,6635
	3	0,9575	0,9838	0,9844	0,9844	0,9844
5	1	0,1953	0,1972	0,1973	0,1974	0,1976
	2	0,3788	0,3872	0,3873	0,3876	0,3878
	3	0,5662	0,5837	0,5843	0,5846	0,5849
	4	0,7504	0,7840	0,7849	0,7850	0,7851
	5	0,9344	0,9835	0,9835	0,9835	0,9835
10	1	0,0945	0,0951	0,0951	0,0951	0,0954
	2	0,1833	0,1853	0,1854	0,1855	0,1858
	3	0,2730	0,2785	0,2786	0,2790	0,2796
	4	0,3653	0,3716	0,3717	0,3721	0,3728
	5	0,4509	0,4668	0,4670	0,4675	0,4682
	6	0,5403	0,5631	0,5634	0,5643	0,5653
	7	0,6316	0,6647	0,6654	0,6667	0,6676
	8	0,7215	0,7666	0,7676	0,7687	0,7695
	9	0,8093	0,8734	0,8752	0,8760	0,8762
	10	0,8990	0,9802	0,9823	0,9824	0,9824

FIGURA 4.12: Gráficos do poder em função do n° de hipóteses falsas dos 5 MCMs empregados na estatística Cramér-von Mises com dados sem censura.

Por meio das Tabelas 4.7, 4.8, 4.9, 4.10, 4.11 e 4.12; e das Figuras 4.7, 4.8, 4.9, 4.10, 4.11 e 4.12; fica claro o que é já bastante afirmado na literatura: o procedimento de Bonferroni apresenta o menor poder em todos os cenários de censura investigados nos testes das duas estatísticas. Já os outros quatro procedimentos apresentaram praticamente

o mesmo comportamento.

Um fato importante a ser notado é que nos testes da estatística de Cramér-von Mises os cinco procedimentos apresentaram maior poder em relação aos testes da estatística *log-rank*. No que diz respeito à censura, os testes realizados com dados sem censura apresentaram maior poder em relação aos dados com censura fixada em 13, que por sua vez apresentaram maior poder em relação aos dados com censura fixada em 11.

4.2 Dados contínuos

Para as simulações realizadas com dados contínuos foram geradas amostras aleatórias de distribuição exponencial. Devido ao elevado custo computacional empregando nos testes as estatísticas *log-rank* e Cramér-von Mises foi analisado apenas o cenário em que aproximadamente 36% dos dados gerados foram censurados.

4.2.1 Controle da FWER

Para averiguar o controle da FWER das cinco estratégias de comparações múltiplas aplicadas na estatística *log-rank* tomamos amostras de tamanho 30, 50, 100, 150, 200, 250 e 300; e para a estatística Cramér-von Mises trabalhamos apenas com amostras de tamanho 30, 50 e 100 devido ao altíssimo tempo computacional gasto. Realizamos 10000 iterações e em cada caso foram geradas 10 populações de distribuição exponencial com parâmetros $\lambda_1, \dots, \lambda_{10} = 0,01$, sendo a primeira população adotada como referência. Fixamos a censura em 100, o que equivale a aproximadamente 36% dos dados censurados.

A Tabela 4.13 mostra os valores da FWER das estratégias de Bonferroni, Holm, Hochberg, Hommel e Rom aplicados na estatística *log-rank* para os diferentes tamanhos de amostras escolhidos. Na Figura 4.13 apresentamos o gráfico com o comportamento da FWER em função do TA para esta mesma estatística.

TABELA 4.13: FWER para os cinco MCMs empregados nos testes da estatística *log-rank* em função dos diferentes TA com censura fixada em 100.

Estratégia	TA=30	TA=50	TA=100	TA=150	TA=200	TA=250	TA=300
Bonferroni	0,03870	0,04200	0,05030	0,04900	0,04890	0,04900	0,04978
Holm	0,03870	0,04200	0,05030	0,04900	0,04890	0,04900	0,04978
Hochberg	0,03870	0,04200	0,05030	0,04900	0,04890	0,04910	0,04985
Hommel	0,03880	0,04200	0,05050	0,04900	0,04910	0,04940	0,04998
Rom	0,03990	0,04320	0,05120	0,04980	0,04990	0,04990	0,05002

FIGURA 4.13: Gráfico do comportamento da FWER dos 5 MCMs empregados na estatística *log-rank*.

Pela Figura 4.13 observamos que com amostras de tamanhos 30 e 50 os cinco métodos mostraram-se bastante conservativos nos testes da estatística *log-rank*. Com TA 150, 200, 250 e 300 o Método de Rom chegou bem próximo ao nível nominal de 5%, enquanto os outros mostraram-se ainda um pouco conservativos atingindo a taxa de 5% apenas com TA igual a 300.

Na Tabela 4.14 e na Figura 4.14 estão listados os valores e o comportamento da FWER dos cinco procedimentos aplicados na estatística Cramér-von Mises para amostras de tamanho 30, 50 e 100.

TABELA 4.14: FWER para os cinco MCMs empregados nos testes da estatística Cramér-von Mises em função dos diferentes tamanhos de amostras (TA) com censura fixada em 100.

Estratégia	TA=30	TA=50	TA=100
Bonferroni	0,04620	0,04830	0,04540
Holm	0,04620	0,04830	0,04540
Hochberg	0,04620	0,04830	0,04540
Hommel	0,04640	0,04840	0,04570
Rom	0,04720	0,04950	0,04630

FIGURA 4.14: Gráfico do comportamento da FWER dos 5 MCMs empregados na estatística Cramér-von Mises.

Pela Figura 4.14 notamos que os 5 procedimentos apresentaram-se conservativos nos testes da estatística Cramér-von Mises com amostras de TA 30 e 100, entretanto com amostra de tamanho 50 a FWER chegou próximo ao nível nominal de 5%. Novamente a estratégia de Rom foi a menos conservativa. Quando comparado aos testes realizados com a estatística *log-rank*, a estatística de Cramér apresentou resultados menos conservativos para TA 30 e 50.

4.2.2 Poder dos testes

Para estimar o poder procedemos das duas maneiras citadas no início deste capítulo. Realizamos em cada estudo 10.000 iterações e trabalhamos com amostras de

tamanho 50.

Para a primeira avaliação do poder foram geradas 10 populações de distribuição exponencial com parâmetros

$$X^{(1)}, \dots, X^{(9)} \sim E(0, 01) \quad \text{e} \quad X^{(10)} \sim E(0, 01 + \delta),$$

em que δ assumiu a sequência de valores 0,002; 0,004; ...; 0,0024; 0,0026, a censura foi fixada em 100 e a primeira população foi tomada como controle.

Na Tabela 4.15 e na Figura 4.15 estão apresentados o poder dos cinco métodos analisados nas estatísticas *log-rank* e Cramér-von Mises em função do incremento de δ .

TABELA 4.15: Poder dos 5 MCMs empregados nas estatísticas log-rank e Cramér-von Mises em função do incremento de δ .

Estatística <i>log-rank</i>					
δ	Bonferroni	Holm	Hochberg	Hommel	Rom
0,002	0,0397	0,0397	0,0399	0,0400	0,0756
0,004	0,0735	0,0737	0,0737	0,0738	0,1097
0,006	0,1906	0,1910	0,1910	0,1917	0,2246
0,008	0,3628	0,3634	0,3635	0,3645	0,3940
0,010	0,5468	0,5474	0,5475	0,5485	0,5679
0,012	0,7074	0,7081	0,7081	0,7085	0,7224
0,014	0,8318	0,8326	0,8326	0,8330	0,8395
0,016	0,9084	0,9087	0,9088	0,9090	0,9135
0,018	0,9524	0,9525	0,9526	0,9529	0,9556
0,020	0,9776	0,9777	0,9777	0,9781	0,9791
0,022	0,9921	0,9921	0,9921	0,9921	0,9926
0,024	0,9958	0,9958	0,9958	0,9958	0,9961
0,026	0,9978	0,9978	0,9978	0,9978	0,9978
Estatística Cramér-von Mises					
δ	Bonferroni	Holm	Hochberg	Hommel	Rom
0,002	0,0315	0,0318	0,0318	0,0319	0,0703
0,004	0,0599	0,0601	0,0601	0,0602	0,1008
0,006	0,1621	0,1629	0,1632	0,1634	0,2014
0,008	0,2853	0,2856	0,2856	0,2867	0,3201
0,010	0,4552	0,4559	0,4559	0,4565	0,4836
0,012	0,6116	0,6120	0,6122	0,6128	0,6338
0,014	0,7381	0,7385	0,7386	0,7396	0,7542
0,016	0,8433	0,8438	0,8439	0,8444	0,8521
0,018	0,8985	0,8987	0,8987	0,8990	0,9048
0,020	0,9418	0,9419	0,9419	0,9423	0,9459
0,022	0,9667	0,9667	0,9667	0,9668	0,9686
0,024	0,9827	0,9828	0,9828	0,9829	0,9834
0,026	0,9898	0,9899	0,9899	0,9899	0,9990

FIGURA 4.15: Gráficos do poder em função do n° de hipóteses falsas dos 5 MCMs empregados na estatística *log-rank*.

Por meio da Tabela 4.15 e da Figura 4.15 podemos ver que o método de Rom apresenta o poder um pouco maior que os demais, principalmente quando δ assume valores pequenos. À medida que δ é incrementado, o poder dos cinco procedimentos ficam bem próximos. Observamos ainda que os testes executados com a estatística *log-rank* apresentam maior poder quando comparado à estatística Cramér-von Mises.

Na segunda investigação do poder trabalhamos com as configurações em que foram geradas 4, 6 e 11 populações para executar respectivamente 3, 5 e 10 testes de hipóteses simultaneamente. As expressões dadas na sequência ilustram a maneira como foram geradas 4 populações para testarmos três hipóteses alternando o número de hipóteses falsas:

$$\begin{aligned} X^{(1)}, X^{(2)}, X^{(3)} &\sim E(\lambda) \quad \text{e} \quad X^{(4)} \sim E(\lambda + \delta), \\ X^{(1)}, X^{(2)} &\sim E(\lambda) \quad \text{e} \quad X^{(3)}, X^{(4)} \sim E(\lambda + \delta), \\ X^{(1)} &\sim E(\lambda) \quad \text{e} \quad X^{(2)}, X^{(3)}, X^{(4)} \sim E(\lambda + \delta), \end{aligned}$$

em que mantivemos $\lambda = 0,01$ e $\delta = 0,015$. De maneira análoga foram criadas as demais populações para realizar os testes com 5 e 10 hipóteses.

As Tabelas 4.16 e 4.17 relatam o poder dos 5 procedimentos dos testes realizados com as estatísticas *log-rank* e Cramér-von Mises, respectivamente. Nas Figuras 4.16 e

4.17 estão os gráficos desse poder quando foram testadas 3, 5 e 10 hipóteses em função do número de hipóteses falsas envolvidas nos testes.

TABELA 4.16: Poder dos 5 MCMs empregados na estatística *log-rank* em função do número de hipóteses falsas.

Número de hipóteses	Número de hipóteses falsas	Estratégia				
		Bonferroni	Holm	Hochberg	Hommel	Rom
3	1	0,3245	0,3301	0,3306	0,3307	0,3307
	2	0,6342	0,6540	0,6552	0,6552	0,6552
	3	0,9439	0,9796	0,9803	0,9803	0,9803
5	1	0,1900	0,1923	0,1923	0,1925	0,1927
	2	0,3723	0,3807	0,3811	0,3816	0,3818
	3	0,5526	0,5741	0,5753	0,5759	0,5762
	4	0,7317	0,7738	0,7749	0,7755	0,7757
	5	0,9160	0,9770	0,9791	0,9791	0,9791
10	1	0,0905	0,0910	0,0910	0,0910	0,0912
	2	0,1767	0,1788	0,1788	0,1790	0,1794
	3	0,2625	0,2686	0,2689	0,2690	0,2698
	4	0,3491	0,3601	0,3602	0,3612	0,3621
	5	0,4343	0,4535	0,4537	0,4548	0,4558
	6	0,5219	0,5502	0,5506	0,5522	0,5535
	7	0,6086	0,6489	0,6501	0,6519	0,6532
	8	0,6950	0,7529	0,7549	0,7571	0,7581
	9	0,7810	0,8621	0,8651	0,8666	0,8670
	10	0,8672	0,9719	0,9758	0,9761	0,9762

FIGURA 4.16: Gráficos do poder em função do n° de hipóteses falsas dos 5 MCMs empregados na estatística *log-rank*.

TABELA 4.17: Poder dos 5 MCMs empregados na estatística Cramér-von Mises em função do número de hipóteses falsas.

Número de hipóteses	Número de hipóteses falsas	Estratégia				
		Bonferroni	Holm	Hochberg	Hommel	Rom
3	1	0,3080	0,3142	0,3152	0,3155	0,3157
	2	0,6015	0,6303	0,6325	0,6327	0,6328
	3	0,8942	0,9529	0,9560	0,9560	0,9561
5	1	0,1789	0,1811	0,1813	0,1817	0,1820
	2	0,3468	0,3569	0,3575	0,3582	0,3588
	3	0,5145	0,5425	0,5441	0,5456	0,5461
	4	0,6832	0,7403	0,7433	0,7443	0,7446
	5	0,8512	0,9449	0,9503	0,9507	0,9508
10	1	0,0823	0,0829	0,0829	0,0829	0,0834
	2	0,1584	0,1610	0,1610	0,1612	0,1619
	3	0,2359	0,2431	0,2433	0,2442	0,2456
	4	0,3159	0,3298	0,3301	0,3313	0,3331
	5	0,3914	0,4152	0,4158	0,4180	0,4205
	6	0,4710	0,5066	0,5078	0,5111	0,5141
	7	0,5483	0,6024	0,6047	0,6087	0,6116
	8	0,6259	0,7019	0,7057	0,7109	0,7134
	9	0,7030	0,8089	0,8166	0,8215	0,8233
	10	0,7783	0,9271	0,9411	0,9427	0,9432

FIGURA 4.17: Gráficos do poder em função do n° de hipóteses falsas dos 5 MCMs empregados na estatística Cramér-von Mises.

Observamos novamente pelas Tabelas 4.17 e 4.16 e por meio das Figuras 4.16 e 4.17 que o procedimento de Bonferroni apresenta poder inferior em relação às demais estratégias. No geral, os procedimentos de Holm, Hochberg, Hommel e Rom apresentaram poder bem semelhantes entre si. Vale notar ainda que os testes empregados com a

estatística *log-rank* resultou em maior poder quando comparado à estatística de Cramér-von Mises.

Capítulo 5

Conclusão e propostas futuras

Nesta dissertação analisamos o desempenho no que diz respeito ao controle da FWER e ao ganho de poder dos procedimentos de comparações múltiplas de Bonferroni clássico e de certos métodos denominados na literatura como seus melhoramentos, dentre eles, o procedimento de Holm, conhecido também como método de Bonferroni de rejeição sequencial; e três procedimentos derivados do método de Simes para realizar inferências sob as hipóteses individuais: Hochberg, Hommel e a recente estratégia proposta por Rom (2013).

Focamos no problema de comparação múltipla de testar um tratamento versus controle empregando nos testes as estatísticas *log-rank* e Cramér-von Mises, ambas com estrutura de dependência. Enfatizamos ainda o trabalho com dados discretos na presença de censura, visto que na área de comparações múltiplas tem se falado pouco desses tipos de dados.

Verificamos que a estatística *log-rank* pertence à classe multivariada totalmente positiva de ordem 2 (MTP_2), fato indispensável para garantir o controle da FWER ao empregar o método de Simes. Para a estatística de Cramér-von Mises ainda não foi possível concluir o mesmo argumento.

Por meio dos estudos de simulação verificamos que em alguns casos os métodos mostraram-se conservativos e em outros extrapolaram um pouco o nível nominal de 5%. Todavia, para tamanhos de amostras grande, o controle da FWER foi satisfatório nos testes das duas estatísticas, tanto para dados contínuos como discretos. Dos cinco MCMs

empregados o que se mostrou menos conservativo e que chegou mais próximo de 5% foi o procedimento de Rom.

Mediante a avaliação do poder o que se pode confirmar claramente é que o procedimento de Bonferroni apresenta menor poder em relação aos demais nos cenários analisados. Entre os procedimentos de Holm, Hochberg, Hommel e Rom não foi notada uma diferença importante.

Como propostas de trabalhos futuros, listamos os seguinte itens:

- Verificar teoricamente se a estatística de Cramér-von Mises satisfaz a condição (MTP_2) ;
- Analisar outros problemas de comparações múltiplas, como por exemplo, testes dois a dois;
- Estudar os métodos de comparações múltiplas que controlam a *false discovery rate* (FDR);
- Analisar alguns métodos que controlam a k -FWER, que é uma versão generalizada da FWER usual. É uma taxa de erro que tem sido introduzida recentemente na literatura.

Apêndice A

Propriedades e resultados assintóticos

Neste apêndice apresentamos alguns resultados que foram utilizados na demonstração de que a estatística log-rank ponderada satisfaz a condição MTP_2 .

Analizamos o caso de J amostras aleatórias independentes $(W_j^p, C_j^p)_{j=1}^{n_p}$, para qualquer $p \in \mathcal{J}$. A fim de manter o controle do comportamento martingale limitante em diferentes amostras introduzimos a filtragem $\mathbb{F} = \{\mathcal{F}_i; i \geq 0\}$ gerada por toda informação disponível em cada categoria da seguinte maneira

$$\mathcal{F}_i := \bigvee_{n_p; p \in \mathcal{J}} \mathcal{F}_i^{n_p}, \quad i \geq 0.$$

Por simplificação, fazemos uso da seguinte notação. Para um dado $p \in \mathcal{J}$,

$$\mathcal{V}_m^p(j) := (V_1^p(j), \dots, V_m^p(j)), \quad \mathcal{R}_m^p(j) := (R_1^p(j), \dots, R_m^p(j)),$$

e

$$\mathcal{R}_m^{C,p}(j) := (R_1^{C,p}(j), \dots, R_m^{C,p}(j)), \quad 1 \leq m \leq N_J; \quad j \geq 1.$$

Introduzimos agora uma outra filtragem que será a base para nossos resultados assintóticos. Esta família de filtragem é cuidadosamente escolhida como descrito na sequência. Para uma dada categoria $j \geq 1$ e $n^* \in \mathbb{N}^J$, definimos a filtragem $\mathcal{G}^{n^*}(j) =$

$\{\mathcal{G}_m^{n^*}(j); 0 \leq m \leq N_J\}$ ao longo das amostras da seguinte maneira

$$\mathcal{G}_m^{n^*}(j) := \sigma\{(V^{n_p}(j), \Delta R^{n^*}(j-1), \mathcal{V}_{m+1}^p(j), \Delta \mathcal{R}_m^p(j), \Delta \mathcal{R}_m^{C,p}(j)), p \in \mathcal{J}\},$$

para qualquer $0 \leq m \leq N_J - 1$ e

$$\mathcal{G}_{N_J}^{n^*}(j) := \sigma\{(V^{n_p}(j), \Delta R^{n^*}(j-1), \mathcal{V}_{N_J}^p(j), \Delta \mathcal{R}_{N_J}^p(j), \Delta \mathcal{R}_{N_J}^{C,p}(j)), p \in \mathcal{J}\},$$

em que definimos $\Delta R^{n^*}(0) = R^{n^*}(0)$ e $\Delta \mathcal{R}_0^p = \Delta \mathcal{R}_0^{C,p} = 0$.

Para qualquer par $q_1 \neq q$ em $\mathcal{J}$ e $n^* = (n_1, \dots, n_J) \in \mathbb{N}^J$, seja $U_{n_{q_1}}^{n_q}(n^*, \cdot) = \{U_{n_{q_1}}^{n_q}(n^*, i); i \geq 1\}$ um processo previsível em relação à filtragem $\mathbb{F}$ satisfazendo os seguintes pressupostos:

(M1) Para cada $(n_q, n_{q_1}) \in \mathbb{N}^2$, $U_{n_{q_1}}^{n_q}(n^*, i)$ é $\mathcal{G}_0^{n^*}(i)$ -mensurável para cada $i \geq 1$;

(M2) Existe $\delta > 0$ tal que

$$\lim_{n^* \rightarrow \infty} n_q \left| \frac{U_{n_{q_1}}^{n_q}(n^*, i)}{V^{n_q}(i)} \right|^{2+\delta} \vee n_{q_1} \left| \frac{U_{n_{q_1}}^{n_q}(n^*, i)}{V^{n_{q_1}}(i)} \right|^{2+\delta} = 0$$

em probabilidade para cada $i \geq 1$;

(M3) Para qualquer $q_2 \in \{q, q_1\}$ e $\ell \geq 1$ existe uma constante $\alpha_{q, q_1}^{q_2}(\ell)$ tal que

$$\frac{|U_{n_{q_1}}^{n_q}(n^*, \ell)|^2}{V^{n_{q_2}}(\ell)} \rightarrow \alpha_{q, q_1}^{q_2}(\ell)$$

em probabilidade com $n^* \rightarrow \infty$.

(M4) Para qualquer $q_2 \in \mathcal{J} (q_2 \neq q, q_2 \neq q_1)$ e $\ell \geq 1$ existe uma constante $\beta_{q_1, q_2}^q(\ell)$ tal que

$$\frac{U_{n_{q_1}}^{n_q}(n^*, \ell) U_{n_{q_2}}^{n_q}(n^*, \ell)}{V^{n_q}(\ell)} \rightarrow \beta_{q_1, q_2}^q(\ell)$$

em probabilidade com $n^* \rightarrow \infty$

A fim de tratar o caso geral, com $3 \leq J \leq \infty$, precisamos introduzir outro par de pressupostos no processo ponderado U . Dado $r \neq k$ em $\mathcal{J}$, seja U um processo ponderado previsível em relação à filtragem $\mathbb{F}$ que satisfaça os seguintes pressupostos:

(H1) Para qualquer $q_1 \neq k$ e $\ell \geq 1$, existe uma constante $\gamma_{q_1}^{k,r}(\ell)$ tal que

$$\frac{U_{n_{q_1}}^{n_k}(n^*, \ell) U_{n_{q_1}}^{n_r}(n^*, \ell)}{V^{n_{q_1}}(\ell)} \rightarrow \gamma_{q_1}^{k,r}(\ell)$$

em probabilidade com $n^* \rightarrow \infty$.

(H2) Para qualquer $q_1 \neq k$ e $\ell \geq 1$, existe uma constante $\eta_{q_1}^{k,r}(\ell)$ tal que

$$\frac{U_{n_{q_1}}^{n_k}(n^*, \ell) U_{n_{q_1}}^{n_r}(n^*, \ell)}{V^{n_{q_1}}(\ell)} \rightarrow \eta_{q_1}^{k,r}(\ell)$$

em probabilidade com $n^* \rightarrow \infty$.

Lema A.1. *Assumimos que os processos ponderados satisfaçam os pressupostos (M1-M4). Assim, para cada $q \in \mathcal{J}$ e $\ell \geq 1$,*

$$\sum_{q_1 \neq q} \left[\frac{U_{n_{q_1}}^{n_q}(n^*, \ell) \mathbb{1}_{\{V^{n_q}(\ell) > 0\}}}{V^{n_q}(\ell)} \Delta Y^{n_q}(\ell) - \frac{U_{n_{q_1}}^{n_q}(n^*, \ell) \mathbb{1}_{\{V^{n_{q_1}}(\ell) > 0\}}}{V^{n_{q_1}}(\ell)} \Delta Y^{n_{q_1}}(\ell) \right] \quad (\text{A.1})$$

converge em distribuição para $N(0, \phi_q^2(\ell))$ com $n^* \rightarrow \infty$. A variância assintótica $\phi_q^2(\ell)$ é dada por

$$\begin{aligned} \phi_q^2(\ell) &= \sum_{q_1 \neq q} \alpha_{q,q_1}^{q_1}(\ell) h^{q_1}(\ell) [1 - h^{q_1}(\ell)] \\ &+ \sum_{q_1 \neq q} \alpha_{q,q_1}^q(\ell) h^q(\ell) [1 - h^q(\ell)] \\ &+ 2 \sum_{q_1, q_2 \in A_q} \beta_{q_1, q_2}^q(\ell) h^q(\ell) [1 - h^q(\ell)], \end{aligned} \quad (\text{A.2})$$

tal que a família de funções $\alpha_{q,q_1}^{q_1}$, α_{q,q_1}^q e β_{q_1,q_2}^q em (A.2) são dadas respectivamente em (M3) e (M4).

Proposição A.1. *Para qualquer par $q_1 \neq q$ em $\mathcal{J}$, assumimos que existem constantes a_{q,q_1} tais que $\lim_{n_q, n_{q_1} \rightarrow \infty} n_{q_1}/n_q = a_{q,q_1}$. Seja f uma função contínua não negativa. Então, os processos ponderados de Harrington e Fleming*

$$U_{n_{q_1}}^{n_q}(n^*, \ell) = \left(\frac{\sum_{i=1}^J n_i}{n_{q_1} n_q} \right)^{1/2} f \left(\prod_{j=0}^{\ell-1} \left(1 - \frac{\Delta R^{n^*}(j)}{V^{n^*}(j)} \right) \right) \left(\frac{V^{n_{q_1}}(\ell) V^{n_q}(\ell)}{V^{n^*}(\ell)} \right),$$

e os processos ponderados de Tarone e Ware

$$U_{n_{q_1}}^{n_q}(n^*, \ell) = \left(\frac{\sum_{i=1}^J n_i}{n_{q_1} n_q} \right)^{1/2} f \left(\frac{V^{n^*}(\ell)}{\sum_{i=1}^J n_i} \right) \left(\frac{V^{n_{q_1}}(\ell) V^{n_q}(\ell)}{V^{n^*}(\ell)} \right),$$

satisfazem os pressupostos **(M1-M4)**.

Demonstração. Precisamos apenas verificar os argumentos para a classe Tarone e Ware, pois os argumentos para Harrington e Fleming são análogos. Fixamos $\ell \geq 1$ e, para abreviar a notação, escrevemos $n = \sum_{i=1}^J n_i$. O pressuposto **(M1)** é óbvio. Se $\delta > 0$, usando a condição de crescimento, a continuidade de f e o fato de que $V^{n^*}(\ell)$ é a soma de distribuições binomiais independentes para obter a seguinte estimativa

$$\begin{aligned} n_q \left| \frac{U_{n_{q_1}}^{n_q}(n^*, i) \mathbb{1}_{\{V^{n_q}(i) > 0\}}}{V^{n_q}(i)} \right|^{2+\delta} &\leq \left(\frac{n}{n_{q_1}} \right)^{1+\delta/2} n_q^{-\delta/2} |f(V^{n^*}(\ell) n^{-1})|^{2+\delta} \\ &\rightarrow 0 \quad \text{com } n^* \rightarrow \infty, \end{aligned}$$

para qualquer $q \neq q_1$ em $\mathcal{J}$. Isto mostra que o pressuposto **(M2)** é satisfeito. Para um dado $q \neq q_1$ em $\mathcal{J}$ e qualquer $q_2 \in \{q, q_1\}$, temos que

$$\frac{|U_{n_{q_1}}^{n_q}(n^*, \ell)|^2}{V^{n_{q_2}}(\ell)} \mathbb{1}_{\{V^{n_{q_2}}(\ell) > 0\}} = \left| f \left(\frac{V^{n^*}(\ell)}{n} \right) \frac{V^{n_{q_1}}(\ell)}{V^{n^*}(\ell)} \right|^2 \frac{V^{n_q}(\ell)}{n_q} \frac{n}{n_{q_1}} \mathbb{1}_{\{V^{n_q}(\ell) > 0\}}. \quad (\text{A.3})$$

Se $q_3 \in \mathcal{J}$ ($q_3 \neq q, q_3 \neq q_1$), escrevemos que

$$\begin{aligned} \frac{U_{n_{q_1}}^{n_q}(n^*, \ell) U_{n_{q_2}}^{n_q}(n^*, \ell)}{V^{n_q}(\ell)} \mathbb{1}_{\{V^{n_q}(\ell) > 0\}} &= \left| f \left(\frac{V^{n^*}(\ell)}{n} \right) \right|^2 \frac{n}{V^{n^*}(\ell)} \left(\frac{n_q n_{q_3}}{n_q n_{q_1}} \right)^{1/2} \\ &\times \left(\frac{V^{n_{q_3}}(\ell) V^{n_q}(\ell)}{n_q n_{q_3}} \right) \left(\frac{V^{n_{q_1}}(\ell)}{V^{n^*}(\ell)} \right) \mathbb{1}_{\{V^{n^*}(\ell) > 0\}}. \quad (\text{A.4}) \end{aligned}$$

As identidades (A.3) e (A.4) nos permite usar novamente a condição de crescimento, a continuidade de f e a propriedade binomial para obter os pressupostos **(M3)** e **(M4)** para concluir a prova. $\square$

Teorema A.1. *Assumimos que um processo ponderado U satisfaça os pressupostos **(M1-***

M4). Assim, temos que

$$\sum_{q_1 \neq q} \left[\int_0^i \frac{U_{n_{q_1}}^{n_q}(\ell)}{V^{n_q}(\ell)} \mathbb{1}_{\{V^{n_q}(\ell) > 0\}} dY^{n_q}(\ell) - \int_0^i \frac{U_{n_{q_1}}^{n_q}(\ell)}{V^{n_{q_1}}(\ell)} \mathbb{1}_{\{V^{n_{q_1}}(\ell) > 0\}} dY^{n_{q_1}}(\ell) \right]$$

converge fracamente para $N\left(0, \sum_{\ell=1}^i \phi_q^2(\ell)\right)$ com $n^* \rightarrow \infty$, para cada $i \geq 1$. Além disso,

$$\sum_{\ell=1}^i \hat{\phi}_{q,n^*}^2(\ell) \rightarrow \sum_{\ell=1}^i \phi_q^2(\ell) \quad (\text{A.5})$$

em probabilidade com $n^* \rightarrow \infty$, para cada $i \geq 1$.

Observação A.1. Para um dado $q \in \mathcal{J}$, denotamos por $Z_q(\cdot)$ o limite fraco da expressão (A.1) no Lema A.1 do seguinte modo

$$Z_q(\ell) := \lim_{n^* \rightarrow \infty} \sum_{m=1}^{N_J} \xi_{m,q}^{n^*}(\ell); \quad \ell \geq 1. \quad (\text{A.6})$$

Estabelecemos ainda

$$W_q(i) := \sum_{\ell=1}^i Z_q(\ell), \quad i \geq 1. \quad (\text{A.7})$$

Observação A.2. Quando $J = 2$, as fórmulas no Teorema A.1 ficam muito simples. Nesse caso, $A_q = \emptyset$ para cada $q \in \mathcal{J}$ e a variância assintótica no Lemma A.1 torna-se

$$\phi_q^2(\ell) = \alpha_{q,q_1}^{q_1}(\ell) h^q(\ell) [1 - h^q(\ell)] + \alpha_{q,q_1}^q(\ell) h^{q_1}(\ell) [1 - h^{q_1}(\ell)], \quad \ell \geq 1.$$

Além disso, $\hat{\phi}_{q,n^*}^2(\ell) \rightarrow \phi_q^2(\ell)$ em probabilidade com $n^* \rightarrow \infty$, em que para um dado $q \in \mathcal{J}$ e n^* ,

$$\hat{\phi}_{q,n^*}^2(\ell) = \frac{|U_{n_{q_1}}^{n_q}(n^*, \ell)|^2}{V^{n_q}(\ell)} \hat{h}^{n_q}(\ell) [1 - \hat{h}^{n_q}(\ell)] + \frac{|U_{n_{q_1}}^{n_q}(n^*, \ell)|^2}{V^{n_{q_1}}(\ell)} \hat{h}^{n_{q_1}}(\ell) [1 - \hat{h}^{n_{q_1}}(\ell)],$$

para cada $\ell \geq 1$.

Lema A.2. Assumimos que um processo ponderado U satisfaça os pressupostos (**M1-M4**) and (**H1-H2**). Dessa forma, para cada $\ell \geq 1$,

$$\left(\sum_{m=1}^{N_J} \xi_{m,1}^{n^*}(\ell), \dots, \sum_{m=1}^{N_J} \xi_{m,J}^{n^*}(\ell) \right) \rightarrow N(\vec{0}, Q(\ell))$$

fracamente com $n^* \rightarrow \infty$, de modo que o operador de covariância $Q(\ell)$ é definido por

$$\langle Q(\ell)a, a \rangle_{\mathbb{R}^J} = \sum_{k=1}^J a_k^2 \phi_k^2(\ell) + 2 \sum_{1 \leq r < k \leq J} a_r a_k \psi(k, r, \ell)$$

para cada $a \in \mathbb{R}^J$.

Proposição A.2. Para qualquer $q_1 \neq q$ em $\mathcal{J}$, assumimos que existem constantes a_{q,q_1} tais que $\lim_{n_q, n_{q_1} \rightarrow \infty} n_{q_1}/n_q = a_{q,q_1}$. Assim, os processos ponderados Harrington e Fleming, e Tarone e Ware satisfazem as suposições **(M1-M4)** e **(H1-H2)**.

Demonstração. Pela proposição A.1, só precisamos verificar **(H1-H2)**. Fixamos $\ell \geq 1$, e para síntese de notação, denotamos $n = \sum_{i=1}^J n_i$. Se $q_1 \neq k \in \mathcal{J}$, escrevemos

$$\begin{aligned} \frac{U_{n_{q_1}}^{n_k}(n^*, \ell) U_{n_{q_1}}^{n_r}(n^*, \ell)}{V_{n_{q_1}}(\ell)} \mathbb{1}_{\{V^{n_{q_1}}(\ell) > 0\}} &= \left| f(V^{n^*}(\ell) n^{-1}) \right|^2 \frac{n}{V^{n^*}(\ell)} \left(\frac{n_r}{n_k} \right)^{1/2} \\ &\times \frac{V^{n_k}(\ell)}{V^{n^*}(\ell)} \frac{V^{n_r}(\ell)}{n_r} \frac{V^{n_{q_1}}(\ell)}{n_{q_1}} \mathbb{1}_{\{V^{n_{q_1}}(\ell) > 0\}} \end{aligned} \quad (\text{A.8})$$

Para um dado $r \neq k$ em $\mathcal{J}$ e $q_1 \neq k$ podemos escrever

$$\begin{aligned} \frac{U_{n_{q_1}}^{n_k}(n^*, \ell) U_{n_k}^{n_r}(n^*, \ell)}{V_{n_k}(\ell)} \mathbb{1}_{\{V^{n_k}(\ell) > 0\}} &= \left| f(V^{n^*}(\ell) n^{-1}) \right|^2 \frac{V^{n_k}(\ell)}{n_k} \frac{V^{n_r}(\ell)}{V^{n^*}(\ell)} \\ &\times \frac{n}{V^{n^*}(\ell)} \left(\frac{n_{q_1}}{n_r} \right)^{1/2} \\ &\times \frac{V^{n_{q_1}}(\ell)}{n_{q_1}} \mathbb{1}_{\{V^{n_k}(\ell) > 0\}}. \end{aligned} \quad (\text{A.9})$$

As expressões (A.8) e (A.9) nos permitem usar a condição de crescimento, a continuidade de f e a propriedade binomial para obtermos os pressupostos **(H1)** e **(H2)**. Como os argumentos para o caso Harrington and Fleming são inteiramente análogos, concluimos a prova. $\square$

Teorema A.2. Assumimos que um processo ponderado U satisfaz **(M1-M4)** e **(H1-H2)**.

Dessa forma, para cada $i \geq 1$,

$$\left(\sum_{\ell=1}^i \sum_{m=1}^{N_J} \xi_{m,1}^{n^*}(\ell), \dots, \sum_{\ell=1}^i \sum_{m=1}^{N_J} \xi_{m,J}^{n^*}(\ell) \right) \rightarrow N(\vec{0}, \Gamma(i)) \quad (\text{A.10})$$

fracamente com $n^* \rightarrow \infty$. Além disso, se denotamos $\hat{\Gamma}(i) = \sum_{\ell=1}^i \hat{Q}(\ell)$ podemos concluir que $\hat{\Gamma}(i) \rightarrow \Gamma(i)$ em probabilidade, em que o operador de covariância $\hat{Q}(\ell)$ é definido por

$$\langle \hat{Q}(\ell)a, a \rangle_{\mathbb{R}^J} = \sum_{k=1}^J a_k^2 \hat{\phi}_k^2(\ell) + 2 \sum_{1 \leq r < k \leq J} a_r a_k \hat{\psi}(k, r, \ell)$$

para cada $a \in \mathbb{R}^J$.

Demonstração. A prova é uma consequência imediata do Lema A.2, em que podemos aplicar os mesmos argumentos do Teorema A.1. De fato, na notação introduzida na Observação A.1 (A.6 e A.7), consideramos

$$\tilde{Z}(\ell) = (Z_1(\ell), \dots, Z_J(\ell)), \quad \ell \geq 1;$$

$$\tilde{W}(i) = (W_1(i), \dots, W_J(i)), \quad i \geq 1.$$

Ao repetir os mesmos argumentos como na prova do Teorema A.1, é fácil verificar que $\tilde{W}$ é um $\mathbb{F}$ -vetor martingale com incrementos independentes, isto é, $\tilde{Z}_i$ e $\tilde{Z}_m$ são variáveis aleatórias independentes com valores em $\mathbb{R}^J$ para cada $i \neq m$. Sob essas condições, através do Lema A.2 concluímos a convergência (A.10). $\square$

Referências

- Curnow, R. & Dunnett, C. (1962). The numerical evaluation of certain multivariate normal integrals. *The Annals of Mathematical Statistics*, **33**(2), 571–579.
- Duchesne, P. & Lafaye De Micheaux, P. (2010). Computing the distribution of quadratic forms: Further comparisons between the Liu-Tang-Zhang approximation and exact methods. *Computational Statistics e Data Analysis*, **54**(4), 858–862.
- Dunnett, C. & Sobel, M. (1955). Approximations to the probability integral and certain percentage points of a multivariate analogue of student's t-distribution. *Biometrika*, **42**(1), 258–260.
- Dunnett, C. W. (1955). A multiple comparison procedure for comparing several treatments with a control. *Journal of the American Statistical Association*, **50**(272), 1096–1121.
- Fisher, R. (1935). The logic of inductive inference. *Journal of the Royal Statistical Society*, **98**(1), 39–82.
- Fleming, T. & Harrington, D. (1991). *Counting processes and survival analysis*, volume 8. Wiley Online Library.
- Harrington, D. & Fleming, T. (1982). A class of rank test procedures for censored survival data. *Biometrika*, **69**(3), 553–566.
- Hochberg, Y. (1988). A sharper bonferroni procedure for multiple tests of significance. *Biometrika*, **75**(4), 800–802.
- Hochberg, Y. & Tamhane, A. (1987). *Multiple Comparison Procedure*. John Wiley, New York.
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, **6**(2), 65–70.
- Hommel, G. (1986). Multiple test procedures for arbitrary dependence structures. *Metrika*, **33**(1), 321–336.
- Hommel, G. (1988). A stagewise rejective multiple test procedure based on a modified bonferroni test. *Biometrika*, **75**(2), 383–386.
- Hommel, G. (1989). A comparison of two modified bonferroni procedures. *Biometrika*, **76**(3), 624–625.

- Hsu, J. (1996). *Multiple Comparison: Theory and methods*. Chapman and Hall, London.
- Karlin, S. (1968). *Total positivity*, volume 1. Stanford University Press.
- Karlin, S. & Rinott, Y. (1980). Classes of orderings of measures and related correlation inequalities. i. multivariate totally positive distributions. *Journal of Multivariate Analysis*, **10**(4), 467–498.
- Leão, D. & Ohashi, A. (2011). On the discrete Cramér-von Mises statistics under random censorship. "*Submitted to the Annal of Statistics*".
- Lehmann, E. E. L. & Romano, J. P. (2005). *Testing statistical hypotheses*. Springer Science+ Business Media.
- Marcus, R., Eric, P. & Gabriel, K. (1976). On closed testing procedures with special reference to ordered analysis of variance. *Biometrika*, **63**(3), 655–660.
- Miller, R. G. (1981). *Simultaneous Statistical Inference*. Springer-Verlag, New York.
- R Development Core Team (2011). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Rom, D. (1990). A sequentially rejective test procedure based on a modified bonferroni inequality. *Biometrika*, **77**(3), 663–665.
- Rom, D. (2013). An improved Hochberg procedure for multiple tests of significance. *British Journal of Mathematical and Statistical Psychology*, **66**(1), 189–196.
- Sarkar, S. (1998). Some probability inequalities for ordered MTP2 random variables: A proof of the Simes conjecture. *The Annals of Statistics*, **26**(2), 494–504.
- Sarkar, S. & Chang, C. (1997). The Simes method for multiple hypothesis testing with positively dependent test statistics. *Journal of the American Statistical Association*, **92**(440), 1601–1608.
- Simes, R. (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika*, **73**(3), 751–754.
- Tarone, R. & Ware, J. (1977). On distribution-free tests for equality of survival distributions. *Biometrika*, **64**(1), 156–160.
- Tukey, J. W. (1953). The problem of multiple comparisons. *The Collected Works of Jonh W. Tukey: Multiple Comparisons, Volume III*, (1999), Edited by H. Braun, CRC Press, Boca Ranton, Florida.
- Wright, S. (1992). Adjusted p-values for simultaneous inference. *Biometrics*, **48**(4), 1005–1013.