UNIVERSIDADE FEDERAL DE SÃO JOÃO DEL REI

Erros envolvidos em testes simultâneos: erro tipo I por comparação ou erro tipo I por experimento?

PROJETO APRESENTADO À PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO, POR EXIGÊNCIA DO EDITAL Nº 004/2020/PROPE, REFERENTE AO PERÍODO DE MARÇO DE 2021 A FEVEREIRO DE 2022.

Orientador: **Ben Dêivide de Oliveira Batista**
ben.deivide@ufsj.edu.br

Ouro Branco - MG

2020

SUMÁRIO

1 INTRODUÇÃO

Em meio a essa pandemia do Covid-19, muitos estudos estão sendo feitos sobre esse vírus, para tentar entendê-lo. Diversas técnicas em análises estatísticas pode ser feitas. Dentre estas, pode-se destacar os procedimentos de comparações múltiplas (PCMs) ou testes simultâneos, dos quais objetivam diferenciar a comparação de dois ou mais grupos de médias, quando há uma real diferença.

Um dos interesse dos PCMs está em verificar um conjunto de parâmetros originados da combinação linear de médias tomadas duas a duas, expressos por $\theta_l = \mu_i$-$\mu_{i'}$, $i \neq i' \in \{1, 2, \ldots, n\}$, para $l = 1$, 2, . . ., N. A coleção desses parâmetros é chamado de família (HOCHBERG; TAMHANE, 1987, p. 5), sendo o número de parâmetros dado por

$$N = \binom{n}{2} = \frac{n(n-1)}{2}, \tag{1}$$

considerando as diferenças das n médias duas a duas.

Cada parâmetro é avaliado por meio de um teste de hipóteses. E cada teste está associado erros de decisão: erro tipo I e erro tipo II. Assim, o desenvolvimento de um PCM é realizado avaliando o seu desempenho por meio do erro tipo I e o poder, sendo este último o complemento do erro tipo II. Essa avaliação se baseia em um processo de simulação, em que nesse estudo será com base no método Monte Carlo. Para a avaliação do erro tipo I, a simulação realizada supõe hipótese nula global (H_0), isto é, a simulação realizada gera amostras de n tratamentos em um experimento da mesma população, portanto, com mesma média populacional. Define-se a hipótese H_0 da seguinte forma,

Definição 1.1 — Hipótese nula global $\mathbf{H_0}$. Seja um fator, de efeito fixo e n níveis, então a hipótese nula global é definida por

$$H_0 \quad : \quad \mu_1 = \mu_2 = \ldots = \mu_n, \tag{2}$$

sendo μ_1, μ_2, . . ., μ_n as n médias das populações. ■

A avaliação do poder do teste é baseada na hipótese H_1 completa, definida por:

Definição 1.2 — Hipótese $\mathbf{H_1}$ completa. Seja um fator, de efeito fixo e n níveis, então a hipótese H_1 completa é definida por

$$H_1 \quad : \quad \mu_1 \neq \mu_2 \neq \ldots \neq \mu_n, \tag{3}$$

sendo μ_1, μ_2, . . ., μ_n as n médias de n populações. ■

A simulação realizada gera amostras de n tratamentos em um experimento de n populações com médias populacionais diferentes. Contudo, na prática essas n médias dos tratamentos não pertencem necessariamente a n populações. O que há na realidade são g grupos de médias pertencentes a g populações, com

$g \leq n$. Nesse caso, a hipótese é expressa numa configuração de médias populacionais em que algumas, mas não todas as médias sejam iguais. Essa hipótese é chamada de hipótese nula parcial. Uma definição formal é apresentada a seguir.

Definição 1.3 — Hipótese nula parcial $\mathbf{H_{0_p}}$. Seja um fator qualquer, de efeito fixo, com n médias e g grupos, em que $g \leq n$, e as p_k médias iguais, para $k = 1, 2, \ldots, g$, estão dispostas no j-ésimo grupo, com $j = 1, 2, \ldots, p_k$, então a hipótese nula parcial é definida por

$$\begin{aligned} H_{0_p} : \mu_{11} = \mu_{12} \ldots = \mu_{1p_1} \neq \ldots \neq \mu_{21} = \mu_{22} \ldots = \mu_{2p_2} \neq \ldots \neq \\ \neq \ldots \neq \mu_{g1} = \mu_{g2} \ldots = \mu_{gp_g}, \end{aligned} \tag{4}$$

sendo $\sum_{k=1}^{g} p_k = n$. ■

Observe que essa definição acaba sendo uma hipótese geral em relação às anteriores, pois se for considerado $g = 1$, nada mais é do que a hipótese nula global. Se $g = n$, tem-se a hipótese H_1 completa. Assim, sob H_{0_p} parcial, dentro do j-ésimo grupo, pode ser avaliado o erro tipo I, e entre grupos, o poder do teste.

De um modo formal, define-se o erro tipo I,

Definição 1.4 — Erro tipo I. Seja H_0 uma hipótese verdadeira. A decisão equivocada de rejeitar H_0, sendo esta verdadeira, é chamada de erro tipo I. O tamanho do erro tipo I é definido pela probabilidade $\alpha = P(\text{Rejeitar } H_0|H_0 \text{ verdadeira})$. ■

Diz-se que quando, em uma inferência, se rejeita uma hipótese, dado que ela é verdadeira, esta é chamada de inferência falso positiva. O ideal para um teste é manter o tamanho do erro tipo I no nível de significância adotado pelo pesquisador, chamado de nível de significância global. Por outro lado, não rejeitar uma hipótese que realmente deveria ser rejeitada, comete-se o erro tipo II.

Definição 1.5 — Erro tipo II. Seja H_0 uma hipótese falsa. A decisão equivocada de não rejeitar H_0, sendo esta falsa, é chamada de erro tipo II. O tamanho do erro tipo II é definido pela probabilidade $\beta = P(\text{Não Rejeitar } H_0|H_0 \text{ falsa})$. ■

Se por um lado, a probabilidade de se incorrer no erro tipo I é conhecida e fixada pelo pesquisador, por outro, a probabilidade do erro tipo II não é conhecida e nem pode ser especificada *a priori*. O que se sabe é que a probabilidade α é inversamente proporcional à probabilidade β. Relacionado ao erro tipo II, uma outra probabilidade que se torna mais fácil de ser entendida é o poder do teste, sendo apresentado a seguir.

Definição 1.6 — Poder do teste. A probabilidade de rejeitar uma hipótese, dado que ela é falsa, denotada por $1 - \beta$, é chamada de poder do teste. ■

A Tabela 1 mostra o resumo dos tipos de erros envolvidos numa tomada de decisão sob a hipótese nula H_0 e nas respectivas probabilidades.

Tabela 1 Tipos de erros e decisões corretas, com suas respectivas probabilidades em um teste de hipótese.

Decisão	**Realidade**	
	H_0 verdadeira	**H_0 falsa**
Não Rejeitar H_0	Decisão correta	Erro Tipo II
	$1-\alpha$	β
Rejeitar H_0	Erro tipo I	Poder
	α	$1-\beta$

Diante dessas considerações, segundo Rao e Swarupchand (2009, p. 66) o termo "comparações múltiplas" refere-se à aplicação de vários testes de significância estatística das diferenças entre as médias (ou proporções ou variâncias, etc) dentro de um grupo. Esses autores ainda definem os procedimentos de comparações múltiplas, apresentados a seguir.

Definição 1.7 — Procedimentos de comparações múltiplas. Os procedimentos estatísticos designados ao controle adequado de efeitos de multiplicidade são chamados de procedimentos de comparações múltiplas. ■

O problema é que se inúmeras hipóteses, e em especial no tipo H_{0_l} : μ_i-$\mu_{i'}$, para $i \neq i' \in \{1, 2, \ldots, n\}$ e $l = 1$, 2, ..., N, são testadas simultaneamente, nos testes tradicionais ocorre um crescimento no erro tipo I. Este é o efeito de multiplicidade, isto é, o crescimento do erro tipo I com o aumento do número de testes. A forma de medir esse crescimento é realizada pela taxa de erro, definida a seguir.

Definição 1.8 — Taxa de erro em inferências simultâneas. Seja um conjunto de hipóteses H_{0_l} verdadeiras e independentes, com $l = 1, 2, \ldots, N$. A probabilidade de pelo menos uma hipótese H_{0_l} ser rejeitada é chamada de taxa de erro, definida por

$$\text{Taxa de erro} = 1 - \prod_{l=1}^{N}[1 - P(\text{Rejeitar } H_{0_l}|H_{0_l} \text{ verdadeira})]. \tag{5}$$

■

Um exemplo para ilustrar essa situação é:

■ **Exemplo 1.1 — Efeito de Multiplicidade de PCMs.** Considere um conjunto de 10 testes de hipóteses verdadeiras e independentes, com probabilidade de erro tipo I igual a $P(\text{erro tipo I}) = \alpha$, sendo $\alpha = 0{,}05$. A taxa de erro é dada por:

$$\begin{aligned} \text{Taxa de erro} &= 1 - \prod_{l=1}^{N}(1-\alpha) \\ &= 1 - (1-\alpha)^N \\ &= 1 - (1 - 0{,}05)^{10} \\ &= 0{,}4013. \end{aligned} \tag{6}$$

Considerando 30 testes de hipóteses verdadeiras, a taxa de erro é $0{,}7854$. Se forem considerados 60 testes, a taxa de erro será $0{,}9539$. Na Figura 1, observa-se assim, que à medida que o número de testes aumenta, a taxa de erro aumenta. Essa taxa de erro nada mais é do que o erro tipo I conjunto de todos os testes.

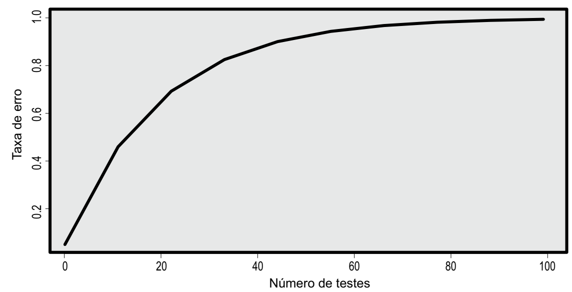

Figura 1 Representação da taxa de erro, em inferências simultâneas, em relação ao número de testes.

■

Percebe-se que na realidade que as N comparações não são independentes, de modo que essa não é a probabilidade exata, mas uma aproximação conservadora (HOCKING, 2003, p. 655). Diz-se então que esta é a probabilidade máxima ou um limite superior de cometer pelo menos um erro tipo I.

Assim, é com base nas taxas de erro tipo I que os procedimentos de comparações múltiplas são desenvolvidos, com o objetivo de controlar adequadamente a probabilidade de pelo menos uma hipótese verdadeira ser rejeitada, isto é, mesmo que com o aumento dos testes de hipóteses não haja também o aumento na taxa de erro.

Outro ponto importante para a avaliação dos PCMs ser baseada no erro tipo I, é que o erro tipo II não pode ser controlado diretamente pelo pesquisador. Segundo Hochberg e Tamhane (1987, p. 10), a medida de probabilidade do erro tipo I é frequentemente escolhida devido à facilidade de ser analisada e controlada.

Nos PCMs existem duas formas de avaliar o erro tipo I. A primeira avaliação é feita verificando-se a probabilidade de rejeitar uma hipótese verdadeira em todas as possíveis combinações dos níveis dos tratamentos tomados dois a dois, sendo conhecida por taxa de erro tipo I por comparação (TEC), do inglês, “comparisonwise error rate” ou “per-comparison error rate” (RAMALHO; FERREIRA; OLIVEIRA, 2005, p. 86). Segundo O’Neill e Wetherill (1971, p. 221) esse tipo de erro é calculado por:

$$\text{TEC} \quad : \quad \frac{\text{N}^{\text{o}} \text{ de comparações rejeitadas indevidamente}}{\text{Número total de comparações}}.$$

A segunda forma é avaliada pela taxa de erro tipo I por experimento (TEE), do inglês “experimentwise error rate” (FERREIRA, 2009, p. 215). Segundo O’Neill e Wetherill (1971, p. 221) esse tipo de erro é calculado por:

$$\text{TEE} : \frac{\text{N}^{\text{o}} \text{ exper. com pelo menos uma hipótese rejeitada indevidamente}}{\text{Número total de experimentos}}.$$

A relação entre esses dois tipos de taxas de erro é relatado por Hochberg e Tamhane (1987, p. 8), dada por

$$TEE = 1 - (1 - TEC)^N, \qquad (7)$$

sendo N o número de combinações a serem feitas, conforme expressão (1).

2 OBJETIVOS

A questão sobre qual das taxas de erros deve ser controlada é motivo de muitas discussões, pois além da relação teórica, há questões filosóficas sobre essa decisão. Assim, este presente trabalho tem o objetivo de apresentar características sobre esses dois tipos de erros, para que o pesquisador ao desenvolver algum procedimento de comparação múltipla, tome a decisão de qual erro deve se basear.

3 METODOLOGIA

As características apresentadas sobre qual dos erros, por comparação ou por experimento, os procedimentos de comparações múltiplas devem ser baseados, o presente estudo tomará essa decisão por meio de evidências científicas e por meio de simulação. Para isso, será levantado todas as discussões sobre esse assunto na literatura. Posteriormente, por meio de alguns testes presentes na literatura, os seus desempenhos serão simulados com base nesses dois tipos de erros, para que assim, se possa verificar qual dos testes realmente consegue controlar o nível de significância global.

4 PLANO DE TRABALHO

Este projeto objetiva o desenvolvimento em um período de doze meses, sendo dividido em quatro etapas:

Primeira etapa

- Revisão bibliográfica sobre as discussões dos tipos de erros utilizados para avaliar o erro tipo I nos procedimentos de comparações múltiplas.

Segunda etapa

- Simulação dos erros tipo I por comparação e por experimento de alguns testes presentes na literatura.

Terceira etapa

- Confronto das informações obtidas na literatura com os da simulação (Primeira e Segunda etapas).

Quarta etapa

- Elaboração de relatório final e de material para apresentação dos resultados obtidos em eventos de divulgação científica.

5 CRONOGRAMA

Na Figura 2 é apresentado o cronograma do projeto, mostrando o período em que cada etapa será realizada.

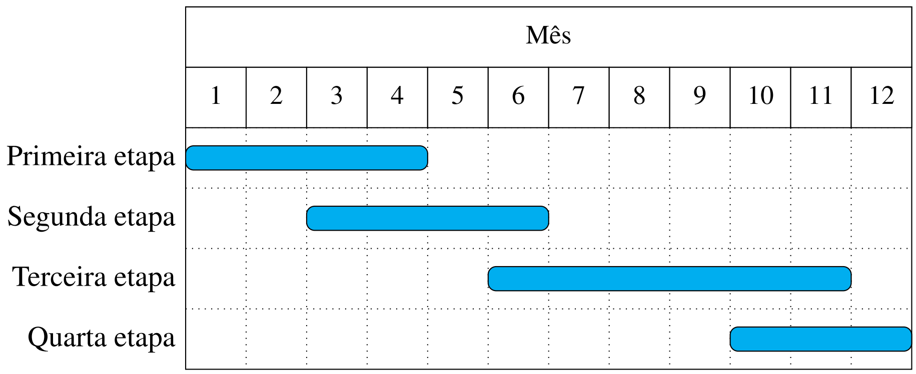

Figura 2 Cronograma de atividades do projeto.

REFERÊNCIAS

FERREIRA, D. F. **Estatística Básica**. 2. ed. Lavras: Editora UFLA, 2009. 664 p.

HOCHBERG, Y.; TAMHANE, A. **Multiple comparison procedures**. New York: Wiley, 1987. 450 p.

HOCKING, R. R. **Methods and Applications of Linear Models**: Regression and the analysis of variance. 2. ed. New Jersey: Wiley, 2003. 741 p.

O'NEILL, R.; WETHERILL, G. B. The present state of multiple comparison methods. **Journal of the Royal Statistical Society**, v. 33, n. 2, p. 218–250, 1971.

RAMALHO, M. A. P.; FERREIRA, D. F.; OLIVEIRA, A. C. de. **Experimentação em genética e melhoramento de plantas**. 2. ed. Lavras: Editora UFLA, 2005. 322 p.

RAO, C. R.; SWARUPCHAND, U. Multiple comparison procedure - a note and a bibliography. **Journal of Statistics**, v. 16, p. 66–109, 2009.