

ANÁLISE E TAXAS DE ERRO DE TESTES DE COMPARAÇÕES MÚLTIPLAS

SILVA, Sunamita Souzaⁱ; SERAPHIN, José Carlosⁱⁱ

Palavras-chave: Simulação – Taxas de erro experimental – Intervalo de confiança exato,

1. INTRODUÇÃO (justificativa e objetivos)

Apesar da existência de uma ampla variedade de textos e artigos sobre procedimentos de comparações múltiplas (STEEL & TORRIE, 1980; PERECIN & BARBOSA, 1988; EINOT & GABRIEL, 1995; SILVA et al, 1999; RAMALHO et al, 2000), ainda há espaço para pesquisas sobre o tema, principalmente quanto ao estabelecimento de critérios e condições para uma recomendação precisa. A possibilidade de utilizar um dos vários testes estatísticos existentes para a comparação de médias é um desafio para o pesquisador, pois a escolha do testes inadequada pode lhe causar dificuldades na interpretação dos resultados ou até mesmo levá-lo a estabelecer conclusões equivocadas. Neste trabalho simulamos e comparamos alguns testes de comparações múltiplas, em várias situações, a partir das taxas de erro , no intuito de poder auxiliar o pesquisador na escolha do teste adequado para cada situação.

2. METODOLOGIA

Os dados utilizados foram gerados por simulação seguindo o modelo matemático inteiramente casualizado.

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij} \quad [1]$$

onde:

y_{ij} = é o valor correspondente a combinação do i - ésimo tratamento na j - ésima repetição .

μ = média geral arbitrada

τ_i = efeito do i - ésimo tratamento (i = 1,2, ... , I)

ε_{ij} = erro experimental aleatório independente, sendo que : $\varepsilon_{ij} \sim N(0, \sigma^2)$.

As simulações foram feitas gerando-se 1000 experimentos para cada uma das combinações de tratamentos (5 e 10), número de repetições (2, 4 e 10), coeficiente de variação (5%, 15%, 20%, 30%, 35% e 45%), para um nível nominal de significância de 1% e 5%. Foi considerada a não existência de efeito do tratamento, ou seja, considerou-se como uma situação de nulidade completa.

Nesta pesquisa avaliamos as taxas de erro por experimento (erro tipo I) dos teste de Tukey, Duncan, LSD e SNK. Para decidir se estas taxas de erro tipo I eram semelhantes aos níveis nominais estabelecidos, calculamos um limite máximo que estes valores poderiam assumir. Para obter esses limites utilizamos o intervalo de confiança exato para proporções (LEEMIS & TRIVED, 1996), que foram calculados através da seguinte expressão:

$$\frac{1}{1 + \frac{n-y+1}{yF_{2y, 2(n-y+1), 1-\alpha/2}}} < p < \frac{1}{1 + \frac{n-y}{(y+1)F_{2(y+1), 2(n-y), \alpha/2}}} \quad [2]$$

onde:

n= nº de casos;

y= nº de sucessos;

F= valor de F referente à probabilidade especificada na cauda direita da distribuição de F.

A simulação, análise dos dados e aplicação dos testes foram realizados usando os procedimentos do programa estatístico **SAS** (Statiscal Analysis System). (SAS Institute Inc, 1985).

3. RESULTADOS E DISCUSSÃO

No que se refere aos resultados das simulações, observando a Tabela 1 nota-se que quando comparamos os testes de Tukey, Duncan, LSD e SNK, a variação do número de tratamento de 5 para 10 provocou nos testes de Duncan e LSD um grande aumento na suas taxas de erro tipo I, tanto para os níveis nominais de significância igual a 5% e 1%, o que difere da conclusão de Bernhardson (1975), onde o aumento no número de tratamento provocou uma queda na taxa de erro tipo I. Já para os testes de Tukey e SNK essa variação no número de tratamentos provocou uma pequena oscilação dentro da taxa de erro esperada. Os limites superiores esperados, encontrados para α igual a 1% e 5% foram 1.8313 % e 6.5391 % respectivamente, e os inferiores foram 0.4806 % e 3.7335 %, calculados usando a expressão 2. Os valores que superaram esses limites não foram considerados semelhantes aos níveis nominais de significância.

Quanto à variação do número de repetições não se observou nenhum padrão, pois, quando observadas isoladamente as taxas de erro permaneceram oscilando em todos os testes, nos dois níveis nominais de significância.

Ainda considerando a Tabela 1 nota-se que os testes de Tukey e SNK deram resultados próximos do esperado, ou seja, os seus resultados estão dentro dos limites do seu nível de significância. Já nos testes LSD e de Duncan chegamos à conclusão que devem ser usados com cautela, visto que apresentaram taxas de erro muito elevadas. No teste LSD esse erro chegou a 48% para 10 tratamentos e 10 repetições, e no teste de Duncan chegou a 37%. Isto discorda dos resultados obtidos por Conagin & Zimmermann (1990) que concluíram que o teste LSD apresentava melhores resultados que o de Tukey.

Nos testes de Tukey e SNK, poucos valores ficaram fora do intervalo de confiança, apenas 2 (para $\alpha=5\%$), para $\alpha=1\%$ nenhum valor foi encontrado fora do intervalo. O contrário acontece com os testes de Duncan e LSD, onde quase todos os valores encontrados ficaram fora dos limites esperados.

Não foi observado efeito do coeficiente de variação sobre as taxas de erro tipo I quando consideradas em uma visão global, mas quando observada a sua variação para o número de tratamento pequeno (no caso 5) observamos que o CV provocou maiores oscilações da taxa de erro tipo I.

Tabela1- Taxa de erro tipo I para os níveis nominais de significância ($\alpha= 5\%$ e $\alpha= 1\%$), em função do número de tratamento, de repetições, coeficientes de variação, para os testes de Tukey, Duncan, LSD e SNK.

Nº de tratamento	Nº de repetições	CV%	Tukey		Duncan		LSD		SNK	
			5%	1%	5%	1%	5%	1%	5%	1%
5	2	5	5.9	1.4	19.3	4.8	21.8	5.9	6.0	1.3
5	2	15	3.8	1.1	18.1	3.1	20.7	3.8	3.6	1.1
5	2	20	7.4	1.7	19.9	5.4	22.9	6.4	7.3	1.6
5	2	30	4.1	0.6	16.6	3.1	19.4	4.3	4.2	0.6
5	2	35	5.1	1.0	16.2	4.1	19.5	5.0	5.0	1.0
5	2	45	5.1	0.8	20.2	3.4	22.7	4.7	5.1	0.8
5	4	5	4.0	1.1	18.3	3.8	25.9	5.8	4.2	1.1
5	4	15	5.6	1.0	18.1	6.2	24.1	6.2	5.6	1.0
5	4	20	6.4	0.9	21.1	4.5	25.0	7.9	6.4	0.9
5	4	30	3.8	0.9	16.0	2.8	23.2	5.5	3.9	0.9
5	4	35	5.2	1.1	18.5	4.0	25.3	6.9	5.5	1.1
5	4	45	6.0	1.4	19.3	5.2	26.0	7.4	6.1	1.5
5	10	5	5.2	1.5	18.3	3.8	27.3	7.3	5.1	1.5
5	10	15	5.2	0.6	17.9	3.5	26.0	6.8	4.5	0.5
5	10	20	4.9	0.6	18.8	4.0	29.1	7.0	5.0	0.6
5	10	30	5.8	1.7	20.3	4.3	30.0	8.1	5.8	1.8
5	10	35	3.6	0.8	19.5	3.0	27.8	6.2	3.7	0.8
5	10	45	5.0	1.0	18.9	3.9	28.0	6.5	5.0	1.0
10	10	5	5.3	1.4	34.1	8.8	47.4	14.6	5.3	1.4
10	10	15	4.3	1.1	37.1	8.9	48.1	13.7	4.1	0.8

4.CONCLUSÃO

A partir dos resultados podemos concluir:

- Os testes que obtiveram melhores resultados foram o teste de Tukey e SNK, e apresentaram taxas de erro tipo I muito semelhantes;
- A variação do CV (coeficiente de variação) não teve grande influência na taxa de erro tipo I;
- Os testes de Duncan e LSD apresentaram altas taxas de erro tipo I, fora dos limites esperados;
- O teste LSD produz taxas de erro maiores que o teste de Duncan,
- Os testes LSD e Duncan devem ser usado com cautela;
- Serão necessárias mais simulações para a obtenção de resultados mais conclusivos;

5. REFERÊNCIAS BIBLIOGRÁFICAS

BERNHARDSON, C.S. Type I error rates when multiple comparison procedures follow a significant F test of ANOVA. **Biometrics**. Washington, 1975. 31(1): 337-340.

CONAGIN, A.; ZIMMERMANN, F.J.P. Seleção de materiais nos trabalhos de melhoramento de plantas. II. Poder discriminativo de “diferentes testes estatísticos”. **Pesq. Agropec. Bras**, Brasília, 1990, 25(10): 1415-1428.

EINOT, I.; GABRIEL, K.R. A study of the power of Several Methods of Multiple Comparisons. **Journal of the American Statistical Association**. September 1995, Volume 70, Number 351, Applications Section.

LEEMIS, L.; TRIVEI, K.S. A comparison of approximate interval estimators for the Bernoulli parameter. **The American Statistician**. Alexandria, 1996. v. 50, n. 1, p.63-68.

PERECIN, D.; BARBOSA, J.C. Uma avaliação de seis procedimentos para comparações múltiplas. **Revista de Matemática e Estatística**. Marília, 1988, 6:95-103.

RAMALHO, M.A.P; FERREIRA, D.F.; OLIVEIRA, A.C. **Experimentação em Genética e Melhoramento de Plantas**. 1. ed. Lavras, 2000. 303p.

SAS Institute Inc. User's Guide: Statistics, 5. ed. Cary NC, 1985

SILVA, E.C.; FERREIRA, D.F.; BEARZOTTI, E. **Avaliação do poder e taxas de erro tipo I do teste de Scott-Knott por meio do método de Monte Carlo**. 1999. Dissertação (Mestrado em Estatística)-Instituto de Ciências Exatas, Universidade Federal de Lavras, Lavras, 1999.

STEEL, R.G.D.; TORRIE, J.H. Principles and Procedures of Statistics. A Biometrical, 1980, App. . McGraw-Hill Book Comp 2 ed. 633p.

¹Bolsista de iniciação científica. Instituto de Matemática e Estatística-IME/UFG, sunamita@mat.grad.ufg.br

² Orientador/Instituto de Matemática Estatística-IME/UFG, seraphin@mat.ufg.br