PODER E EFICIÊNCIA DOS DIFERENTES TESTES ESTATÍSTICOS PARA COMPARAÇÕES MÚLTIPLAS

Armando Conagin[1], Décio Barbin[2]

RESUMO

Para avaliar o poder de dezesseis testes estatísticos de forma a decidir sobre o mais adequado, foram gerados, pelo método de Monte Carlo, mil setecentos e trinta e dois experimentos com 3, 4, 5, 6, 8, e 15 repetições para 8, 12, 13 e 25 tratamentos, em blocos ao acaso, com coeficientes de variação de 5%, 10%, 15%, e 20%. A probabilidade de erro tipo I, adotada foi de 0,05.

Pelos resultados obtidos, verifica-se que, para todos os testes, o poder dos testes diminui à medida que a diferença real entre as duas médias, diminui. É possível verificar que os testes de t de Student unilateral, t de Student bilateral e o teste de Waller-Duncan, exibem um poder maior que os demais. Seguem-se pela ordem, o teste de Duncan, o de Dunnett unilateral, o de Sidak modificado, os testes de Bonferroni modificados BM_1 e BM_2, o de Dunnett bilateral, o SNK, o REGWF, o REGWQ, o SMM o de Tukey, o de Sidak original e o de Bonferroni original.

Foi avaliada para cada diferença estudada, a eficiência dos vários testes em relação ao teste de t unilateral.

Com relação ao erro do tipo I, por experimento e do erro I, tipo MEER os resultados evidenciam uma grande perda de eficiência dos testes

[1] Engenheiro-Agrônomo, Estatístico Aplicado, aposentado do Instituto Agronômico, Campinas, S.P., Brasil.
[2] Professor Titular do Departamento de Ciências Exatas, Escola Superior de Agricultura "Luiz de Queiroz", ESALQ, Universidade de São Paulo, Piracicaba, S.P., Brasil.

devido serem mais exigentes; as perdas de eficiência dos testes aumentam, sensivelmente, para diferenças gradualmente menores, o que leva a alertar os pesquisadores para uma escolha dos testes mais eficientes.

Palavras-Chave: Comparações Múltiplas, Testes Estatísticos, Tipos de Erro I, Método de Monte Carlo, Poder dos Testes, Eficiência dos Testes.

POWER AND EFFICIENCY OF DIFFERENT STATISTICAL TESTS FOR MULTIPLE COMPARISONS

ABSTRACT

To evaluate the power of sixteen statistical tests were simulated by Monte Carlo method one thousand seven hundreds and thirty-two experiments with 3, 4, 5, 6, 8 and 15 repetitions with 8, 9, 12, 13 and 25 treatments, in randomized blocks with coefficient of variation of 5%, 10%, 15%, and 20%. The probability of the error term of type I, was 0,05.

It was possible to verify that, for all tests, the respective power will be greater when the differences tested, increased. The differences of treatments against the control were 30%, 20%, 15%, 10% and 5%. The power of the tests decreased and differentiates more and more when the differences between the two means, decreased. It is possible to verify that the tests of Student unilateral, Student bilateral and Waller-Duncan exhibet greater power than the remaining tests, it followed, in order, that the Duncan's test is followed by Dunnett's unilateral, Modified Sidaks, Modified Bonferroni's BM_1 and BM_2, Dunnett's bilateral, SNK, REGWF, REGWQ, SMM, Tukey's, Sidak's original and Bonferroni's original.

It was evaluated the efficiency of the various tests in relation to Student's unilateral, for each difference. The results clearly, evidence that great loss of efficiency occurs in all tests of experimentwise types, the loss of efficiency increases when the difference between means gradually decrease.

Key words – Multiple Comparisons, Statistical Tests, Types of Error I, Monte-Carlo Method, Power of Tests, Efficiency.

INTRODUÇÃO

Nas ciências experimentais comprovam-se as hipóteses formuladas, através dos resultados obtidos em experimentos, com o auxílio da análise estatística e dos testes estatísticos.

Os experimentos são efetuados utilizando um esquema experimental ("design"), que engloba os fatores supostamente relevantes para a elucidação das hipóteses efetuadas. Para maior eficiência os esquemas utilizados devem levar em consideração aspectos importantes: escolha dos tratamentos, escolha das dosagens dos fatores estudados, existência de número adequado de repetições, locação dos tratamentos em blocos, caso haja necessidade de controle local, sorteio dos tratamentos nos blocos de forma a assegurar a aleatoriedade e uma boa distribuição dos mesmos, adoção de técnica experimental esmerada na condução do experimento, precaução na obtenção e registro dos dados etc. Terminado o experimento, a interpretação dos fenômenos pesquisados e a prova das hipóteses feitas é conseguida com o auxílio da análise da variância dos dados e com a aplicação dos testes estatísticos apropriados. Todos esses aspectos devem ser levados em consideração para que as respostas obtidas possam bem representar a situação real da expressão dos fenômenos pesquisados de forma a que se possa esperar repetibilidade dos resultados em situações semelhantes.

A análise estatística a ser feita deve estar de acordo com o delineamento experimental adotado já que uma análise inadequada, quase sempre leva a resultados enganosos.

A interpretação dos resultados é feita pela comparação das respostas obtidas através da aplicação de testes de comparação de médias e/ou de testes de contrastes.

A metodologia utilizada para a escolha da região de rejeição ou não da hipótese é feita de tal forma que, assumindo a não existência de diferenças entre os tratamentos (hipótese nula global), escolhe-se uma área de distribuição da estatística em questão (teste de t, teste de Tukey etc) com probabilidade especificada ($\alpha = 0{,}05$, $\alpha = 0{,}01$ etc), área essa que abrange os resultados do teste a ser efetuado, que têm pequena probabilidade de acontecer se a hipótese nula for verdadeira mas, que têm grande probabilidade de acontecer se a(s) hipótese(s) alternativa(s) (existem diferenças entre as médias dos tratamentos comparados) forem verdadeiras. Essa área de probabilidade $\alpha = 0{,}05$, $\alpha = 0{,}01$ etc é demarcada pelo valor crítico do teste, por t_c no caso da distribuição do teste de t, ou por qc no caso da distribuição do teste da amplitude studentizada (teste de Tukey). Todos os valores de t obtidos pelo teste que pertençam à região com probabilidade $\alpha = 0{,}05$ ou $\alpha = 0{,}01$ etc, rejeitam a hipótese nula H_0 e nos levam a adotar a hipótese alternativa H_a.

Os estatísticos desenvolveram mais de uma dezena de testes estatísticos, cada um deles com a sua formulação lógica, baseada em determinada função de distribuição.

A área adotada para rejeição da hipótese nula (verdadeira) é a área conhecida como área do erro I ou de 1ª espécie. Ela é escolhida de tal forma que seja mínimo o erro de 2ª espécie (erro ß) que consiste na probabilidade de adoção de uma hipótese que é falsa. Por exemplo, supor que exista uma diferença real entre duas médias (a hipótese H_a é verdadeira) e o resultado do experimento proporciona um valor do teste, menor que o valor crítico adotado, supor tc. Então H_0 não vai ser rejeitada e a hipótese nula permanece apesar de falsa. Estamos cometendo um erro de 2ª espécie, ß (adotar uma hipótese que é falsa).

Em pesquisas de saúde pública, com vistas à adoção de um novo medicamento, para ser aprovado para comercialização, experimentos têm que ser feitos para comparar o novo medicamento com o atualmente recomendado. Na adoção de um novo híbrido de milho, por exemplo, este vai ser comparado através de experimentos comprobatórios, com o híbrido ou cultivar em uso pelos agricultores. Esse procedimento também é seguido nas áreas da veterinária, engenharia de produção, biologia, indústria química etc. Portanto, comparações com uma testemunha representam um lugar comum nas pesquisas experimentais. Este trabalho tem como objetivo avaliar o poder de dezesseis testes estatísticos, sob diversas situações.

MATERIAL E MÉTODO

A avaliação do poder de diferentes testes estatísticos teve início com Gabriel (1964), O Neill e Wetheril (1971), Boardman e Moffitt (1971) e vários outros. Além da comparação de médias procuraram avaliar os tipos de erro I por comparação e por experimento. Os tipos de erro I por experimento estão classificados em erro do tipo I por experimento em condições da hipótese nula generalizada, erro tipo I por experimento em condições da hipótese nula parcial e o tipo MEER (SAS, 1990) que avalia os erros do tipo I considerando as duas propriedades.

As pesquisas de Boardman e Moffitt (1971) evidenciaram que, com relação ao erro I por comparação o teste de t de Student para número de tratamentos variando de 2 a 11 em condições de H_0 verdadeira, manteve a taxa do erro I ao redor do valor nominal do erro I adotado: $\alpha = 0,05$; o teste de Duncan apresentou valores decrescentes de 0,05 para t=2 diferenças iguais até 0,025 para t=11; os testes SNK, de Tukey e de Scheffé.incluídos na pesquisa apresentaram valores iniciais de 0,05 decrescendo rapidamente para valores 0,01 e menores (t =11).

Com relação ao erro I por experimento, verificaram que a taxa de erro I, para o teste de t, adotado $\alpha = 0,05$, cresceu linearmente quando o número de tratamentos iguais cresceu de 2 para 11 atingindo valor próximo a 0,60; para Duncan o mesmo aconteceu, porém, com menos intensidade crescendo de 0,05 para 0,35. Os outros testes mantiveram α =0,05 (de 2 a 11).

Também Bernardson (1975) pesquisando nas mesmas linhas obteve resultados concordantes com os de Boardman e Moffitt (1971).

No Brasil, Perecin e Barbosa (1988), Conagin (1998), Conagin (1999), Conagin, Igue e Nagai (1999), Conagin (2001), Conagin e Pimentel-Gomes (2004) e Conagin e Barbin (2006) procuraram avaliar o comportamento dos diferentes testes estatísticos para diferentes médias, número de tratamentos, número e de repetições e diferentes coeficientes de variação.

Perecin e Barbosa (1988) mostraram que, para diferenças reais entre médias de $2\sigma_{\bar{x}}$, $4\sigma_{\bar{x}}$, $6\sigma_{\bar{x}}$ e $8\sigma_{\bar{x}}$ para 4 repetições e número variável de tratamentos, os testes de t, SNK, Duncan, Tukey e Waller-Duncan exibiram o poder dos testes respectivos crescendo com o aumento das diferenças entre médias e também com o número de tratamentos (pouca influência); mostraram que o teste de Tukey apresentou um poder de teste bem inferior aos demais.

No trabalho atual os autores ampliaram o leque das pesquisas efetuadas anteriormente, incluindo situações mais diversas o que possibilitou, pela inclusão de uma maior gama de testes, avaliar, não só o poder de cada teste, como a eficiência dos mesmos em relação ao teste de t de Student, unilateral. Apresentaram o poder e a eficiência dos testes de Borferroni Modificados e de Sidak Modificados.

Convém realçar as diferenças entre o erro I por comparação e o erro I por experimento. O erro I, por comparação é o erro que se comete quando a hipótese nula geral é verdadeira e a diferença posta em prova se revela significativa; se efetuam N comparações e se em n comparações se rejeita a hipótese nula (verdadeira), o erro por comparação é dado por n/N.

Erro I, por experimento é aquele que se comete quando do total de N experimentos efetuados, em a experimentos se rejeita falsamente a hipótese nula (verdadeira), uma ou mais vezes em cada experimento. O erro tipo I, por experimento é então dado por a/N. Dos erros por experimentos existem três subtipos Erro Tipo I, por experimento em condições de hipótese nula generalizada H_0: $\tau_1 = \tau_2 = \tau_3 = \ldots = \tau_t$, erro por experimento em condições de hipótese nula parcial, isto é, havendo t tratamentos a hipótese H_0 especifica, $H_0^{'}: \tau_1 = \tau_2 = \tau_3 = \ldots = \tau_\ell$, $\ell < t$. O terceiro subtipo, chamado MEER (SAS, 1990) é o tipo mais geral e engloba os dois anteriores.

No teste de t, de Sstudent, o erro que se comete é do tipo Erro I por comparação; O teste SNK é do tipo erro I por experimento.

Os testes de Tukey, Bonferroni, SMM, Sidak, Dunnett, REGWF, REGWQ são do tipo MEER.

O teste de Dumcan apresenta erro tipo I, por experimento de valores crescentes quando a amplitude dos dois tratamentos comparados, cresce. O teste de Waller-Duncan tem um fundamento diferente dos anteriores e é conhecido como t bayesiano.

TESTE DE BONFERRONI MODIFICADO BM_1

O teste de Bonferroni original procura fixar em α o valor do erro tipo I para o conjunto de comparações; para cada comparação sendo α o valor global, e só existem k comparações a serem feitas o erro adotado para cada uma delas é $\alpha_\beta = \alpha/k$. Se todas as comparações forem feitas, duas a duas, então $k = C_t^2 = t(t-1)/2$. Se forem feitas, unicamente, as comparações dos t-1 tratamentos com um controle, então k = t-1.

Para o teste modificado adota-se α' = α (1+P), para o erro global

$\alpha_{BM_1} = \frac{\alpha'}{k} = \frac{\alpha}{k}(1+P)$; o valor P é obtido da seguinte maneira. Se no experimento F_0 = QMTrat/QMresíduo for significativo, isto é, se $F_0 > F_c$ em que F_c é o valor de F crítico (obtido nas tabelas de F para 0,05 ou 0,01), então existe uma ou mais diferenças entre médias, significativas, já que H_0: $\tau_1 = \tau_2 = \tau_3 = \ldots = \tau_t$ vai ser rejeitada. O número de diferenças significativas e o valor das diferenças estão contribuindo, para que Fo > Fc. A hipótese alternativa, Ha, terá algumas diferenças $\tau_i = \tau_t$ diferentes de zero.

Pode-se avaliar $P(F_0)$ pela distribuição não central de F, sendo $\hat{\lambda}$ (valor da não centralidade da distribuição de F) a qual é estimada como a seguir:

$$\hat{\lambda} = (F_0 - 1)(t - 1)$$

Conagin (1999).

A avaliação de P (F_0) é obtida no SAS (1990) a partir do procedimento Prob Functions. Este calcula:

$$P(F_0) = P_r\ (F_0, (t\text{-}1), (t\text{-}1)\ (r\ \text{-}1), \hat{\lambda})\ \text{e}$$

$$P(F_c) = P_r\ (F_c, (t\text{-}1), (t\text{-}1)\ (r\ \text{-}1), \hat{\lambda}).$$

Calculado P (F_0) e $P(F_c)$, define-se:

$$P = P(F_0) - P(F_c) \quad \text{(Conagin, 1998)}$$

O valor P representa a área da função

$$F(F_i, (t-1), (t-1)(r-1), \hat{\lambda})$$

que abrange todos os resultados de F maiores que F_c (valores que rejeitam H_0) e menores que o valor F_0 obtido na análise. Quanto maior F_0 e sendo F_c fixo no caso, maior será P, que é então uma medida proporcional à grandeza e ao número de diferenças significativas. O valor P está incluído no intervalo $0 \leq P \leq 1$ o que indica que α estará entre α e 2α.

Adota-se, portanto para valor de α_{BM1}

$$\alpha_{BM1} = [\alpha(1+P]/k.$$

Quando $\hat{\lambda}$ é maior que 100, Prob F não calcula o valor de $P(F_0)$ e $P(F_c)$. Para este caso Conagin (2001) obteve tabelas para avaliar as probabilidades para diferentes valores de t, r e F_0/F_c.

Calculando α_{BM1} procura-se na tabela da distribuição de t para graus de liberdade (r-1)(t-1) no caso, e probabilidade do erro I α_{BM1} o valor de t.

Feito o cálculo de

$$t = (\overline{x}_i - \overline{x}_l) / \sqrt{2s^2 / r}$$

se $r_1 = r_2 = \ldots r_t = r$, e, se t obtido for maior que o da tabela o resultado é significativo. Se $r_i \neq r_l$ então em vez de 2/r usa-se $\frac{1}{r_i} + \frac{1}{r_l}$. O resultado, se significativo, indica que há diferença entre as duas médias comparadas pelo teste de Bonferroni Modificado com erro tipo I com probabilidade α_{BM_1}. O valor F_0 engloba o número a de reais diferenças existentes, mas impossível de ser avaliada diretamente. Porém, sabe-se que F_0 as engloba e, portanto, P sendo proporcional ao valor de F_0 já que F_c é fixo, pode-se utilizar P, como uma medida que corrige a situação, pois é uma medida proporcional ao número de diferenças reais, alguma delas de valor menor que não podem ser detectadas pelo teste, devido à grandeza de erro experimental. Conagin (1999) ampliou a pesquisa sobre BM_1 para o caso da hipótese nula parcial.

TESTE DE BONFERRONI MODIFICADO BM_2

Neste teste os autores se propõem a estimar o valor a (número de diferenças reais) efetuando o cálculo da seguinte maneira: feita a análise da variância e se obtiver $F_0 > F_c$, isto é, se F_0 é significativo, propõe-se obter uma estimativa â de a efetuando uma avaliação do numero â pelo calculo do número de diferenças significativas pelo teste de t usando α% como erro por comparação. Se existirem â diferenças significativas, então, H_0: $\tau_1 = \tau_2 = \tau_3 = \ldots = \tau_t$ é rejeitada e se tem H_a (pelo menos uma ou mais $\tau_i \neq \tau_m$ onde o t utilizado é o t unilateral já que as comparações são quase sempre efetuadas contra uma testemunha $\tau_i > \tau_c$), como foi feito nos estudos apresentados.

Sendo k o número de comparações a serem feitas, então para cada comparação:

$$\alpha_{BM2} = \alpha/(k-\hat{a})$$

No caso geral k = $C_t^2 = t(t-1)/2$; no caso de comparação dos t=1 tratamentos com o controle, k = t-1.

O teste de Bonferroni Modificado BM_2 adota daqui para frente, buscar o *t* na tabela com procedimento idêntico ao efetuado para BM_1, porém com o valor da probabilidade do erro I α_{BM2} no lugar de α_{BM1}.

As diferenças significativas obtidas são as do teste BM_2.

TESTE DE SIDAK MODIFICADO

O teste de Sidak original utiliza como probabilidade do erro I o valor α_{S_i} para cada comparação, o qual é avaliado da seguinte maneira:

$$\alpha_{S_i} = 1-(1-\alpha)^{1/k}$$

em que α é a probabilidade do erro I para todas as comparações, (normalmente α = 0,05 ou α = 0,01), k é o numero de comparações feitas, em caso de todas as comparações k = C_t^2 e no caso de t-1 comparações com o controle k = t-1.

Para obter o teste de Sidak modificado, faz-se a análise da variância e se $F_0>F_c$, estima-se a pelo valor â pelo cálculo do número de diferenças significativas usando o teste de t de Student. Se existirem â diferenças significativas, o erro para cada teste de diferença será:

$$\alpha_{S_iM} = 1-(1-\alpha)^{1/k-\hat{a}}$$ (Conagin e Barbin, 2006).

Com o valor da probabilidade erro α_{S_iM} vai-se, à semelhança dos testes modificados anteriores, buscar o valor t_c na tabela de t para (r-1) (t-1) graus de liberdade e p = α_{S_iM}. Adota-se o valor de t obtido e se

$$\text{Dif} > t\sqrt{2s^2/r}$$

a diferença das duas médias será significativa pelo teste de Sidak Modificado.

Como pequenas diferenças reais não são detectadas devido à grandeza do erro experimental, â < a e então o valor de t utilizado é maior que o responsável pelo teste se a fosse detectado e sendo assim, tanto BM_2 como SiM são testes conservadores.

Maiores detalhes sobre estes testes encontram-se em Conagin e Barbin (2006).

RESULTADOS E DISCUSSÃO

Poder dos Testes

A avaliação do poder dos testes obtidos para diferentes situações (quadros 1 a 7) permite efetuar as seguintes observações:

Quando o coeficiente de variação é pequeno CV = 5% e CV = 10%, para diferenças de 30%, 20% o poder dos diferentes testes é sempre alto e praticamente equivalentes. A medida que as diferenças decrescem em magnitude, o poder dos diferentes testes vai se diferenciando nitidamente, sendo que o poder para os casos de r = 3, r = 4 e r = 5, são sempre inferiores aos respectivos, quando r = 6, r = 8 e r = 15.

Esse panorama se repete, para CV = 15% e 20% sendo que os testes apresentam um poder do teste bem inferior aos encontrados nas situações em que CV = 5% e CV = 10%.

Nos casos em que o teste de Student unilateral foi incluído na pesquisa (quadros 5, 6 e 7), verifica-se o seguinte:

O poder do teste t unilateral foi ligeiramente superior ao de Waller-Duncan e t bilateral, e de Duncan, sendo que a superioridade se acentua para diferenças gradualmente menores. De uma maneira geral esses 4 testes apresentam um poder do teste superior aos dos testes de Dunnett unilateral, Sidak modicado, Bonferroni Modificados: BM_1 e BM_2, Dunnett bilateral, REGWF, REGWQ, Tukey, SHM, Sidak original e Bonfenoni original. Os testes modificados têm uma performance semelhante ao de Dunnett unilateral, sendo este último o teste MEER mais eficiente para a comparação de tratamento com um controle.

Os testes modificados foram desenvolvidos pelos autores. O teste BM_1 foi apresentado por Conagin (1998) e ampliado por Conagin (2001) e, ainda, rediscutido, por Conagin e Pimentel Gomes (2004). Os testes de Bonfenoni BM_2 e de Sidak (SiM) modificados foram apresentados por Conagin e Barbin (2006).

Esses testes modificados apresentam um poder de teste equivalente ao de Dunnett e superior aos testes SNK, REGWF, REGWQ, Tukey, Sidak e Bonfenoni originais.

Eficiência dos Diferentes Testes Estatísticos

É importante avaliar a eficiência dos diferentes testes estatísticos, pois, é de interesse do pesquisador utilizar os mais eficientes.

Examinando os quadros 8, 9 e 10, vê-se que, para diferenças de 20% e maiores, os testes apresentam pouca ou nenhuma diferença de eficiência; entretanto, para diferenças gradualmente menores as diferenças de eficiências entre os vários testes se acentua e é do interesse do pesquisador escolher o teste que seja o mais eficiente a partir das diferenças menores.

Pode-se ver nos quadros (8 a 10) que a eficiência dos testes aumenta quando cresce o número de repetições.

Cochran e Cox (1957) desenvolveram a partir da fórmula

$$t = (\overline{x}_i - \overline{x}_j)/\left(s\sqrt{2/r}\right)$$

uma outra que permite ao pesquisador calcular o número de repetições que assegura com alta probabilidade a obtenção de diferenças significativas em 80%, 90%, 95% dos casos, em futuros experimentos.

Para isso, adotam um valor de $t = t_1 + t_2$ em que t_1 é o valor crítico para H_0, com $\alpha = 0{,}05$ ou $\alpha = 0{,}01$. O valor t_2 é obtido a partir da distribuição de t da hipótese alternativa com a função de poder de 80%, 90%, 95% etc. Assim: $t = t_1 + t_2 = (\overline{x}_i - \overline{x}_j)/\left(s\sqrt{2/r}\right)$

que é equivalente à formula: $t_1 + t_2 = dif\%/\left((CV\%)\sqrt{2/r}\right)$

e daí $\sqrt{r} = [\sqrt{2}(t_1 + t_2)(CV\%)]/dif\%$.

Para 80% de probabilidade de obter diferenças maiores ou iguais à dif%, tem-se que, para CV = 5% e dif = 5%, CV = 10% e dif = 10%, CV = 15% e Dif = 15% etc, são necessárias 13 repetições para um único experimento e 7 repetições para 2 experimentos, analisados conjuntamente. Pelo exame dos Quadros 1 a 4 é possível verificar que para CV% = dif% e r

= 15, o número de diferenças significativas se distribui ao redor do valor 80, obtido em porcentagem, concordando com os estudos de Cochran e Cox.

Como conclusão poderemos adicionar o seguinte: cabe ao pesquisador decidir o tipo de teste a ser adotado e o número de repetições dos futuros experimentos, pois o número de repetições e a adoção do teste mais poderoso 'tornará' grandemente provável a obtenção de diferenças significativas e a comprovação das hipóteses formuladas, contribuindo para uma mais rápida solução para os problemas pesquisados.

REFERÊNCIAS BIBLIOGRÁFICAS

BERNARDSON, C. S. 1975. Type Erros Rates When Multiple Comparison procedures Fllours a Significant test of Amovs. **Biometrics**, 3133-340.

BOARDMAN, T.J.; MOFFITT, D.R., 1971. Graphical Monte Carlo Type Y Erros Rates, for Multiple Comparison procedures. **Biometrics**, 27:738-744.

COCHRAN, W.G.; COX, G.M. 1957. **Experimental Design**. John Wiley, Nova Iorque.

CONAGIN, A. 1998. Discriminative Power of Modified Bonferroni Test. **Revista de Agricultura**. 73: 31-46.

CONAGIN, A., 1999. Discriminative Power of The Modified Bonfenoni's Test Under General and Partial Null Hypotheses, **Revista de Agricultura**, 74: 117-126.

CONAGIN, A., 2001. Tables for The Calculation of The probability to be used in The Modified Bonfenoni Test. **Revista de Agricultura**, 76(1): 71-83.

CONAGIN, A., IGUE, I, NAGAI, V. 1999. Poder Discriminativo de Differentes Testes de Médias. Campinas, Instituto Agronômico. **Boletim Cientifico**, 44.

CONAGIN A.; PIMENTEL-GOMES, F. 2004. Escolha adequada dos testes estatísticos para comparações múltiplas. **Revista de Agricultura**. 79:288-295.

CONAGIN, A.; BARBIN, D., 2006, Bonferroni's and Sidak's modified tests. **Scientia Agricola**, 63(1): 70-76.

GABRIEL, K.R. A procedure for treating the homogeneity of all set of means in analysis of variance. **Biometrics**, v. 20, p. 459-477, 1964.

O'NEILL, R.; Wetheril, G.B. 1971. The present state of multiple comparison methods. **Journal of the Royal Statistical Society**, 33: 218-250.

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

STATISTICAL ANALYSIS SYSTEM INSTITUTE – SAS/STAT. 1990. **Users Guide** 4 ed. Carj U.S.A. 890p.

QUADRO 1 - Poder dos diferentes testes estatísticos expresso por porcentagem resultantes da simulação de n=100 e n=33 experimentos com as diferenças reais de 30%, 20%, 15%, 10% e 5%, abrangendo 13 tratamentos com 5 e 15 repetições e coeficiente de variação de CV=5%.

Testes	r = 5, n = 100 experimentos					r = 15, n = 33 experimentos				
	30%	20%	15%	10%	5%	30%	20%	15%	10%	5%
Waller	100	100	100	94	42	100	100	100	100	82
T bilateral	100	100	100	89	36	100	100	100	100	82
Duncan	100	100	99	88	36	100	100	100	100	82
SNK	100	100	96	82	36	100	100	100	100	82
REGWF	100	100	93	62	9	100	100	100	100	48
REGWQ	100	100	94	58	3	100	100	100	100	48
Tukey	100	100	86	38	3	100	100	100	100	30
SMM	100	100	83	32	2	100	100	100	100	27
Sidak	100	100	81	32	2	100	100	100	100	27
Bonferroni	100	100	81	32	2	100	100	100	100	27
Dunnett	100	100	98	73	15	100	100	100	100	55

QUADRO 2 - Poder dos diferentes testes estatísticos expressos em porcentagem resultantes da simulação de n=100 e n=33 experimentos com as diferenças reais de 30%, 20%, 15%, 10% e 5%, abrangendo 13 tratamentos com r = 5 e r = 15 repetições e coeficientes de variações CV=10%.

Testes	r = 5, n = 100 experimentos					r = 15, n = 33				
	30%	20%	15%	10%	5%	30%	20%	15%	10%	5%
Waller	100	88	65	46	12	100	100	97	76	21
T bilateral	100	85	58	32	8	100	100	91	70	18
Duncan	100	79	50	27	6	100	100	91	67	12
SNK	96	64	34	20	1	100	100	88	61	12
REGWF	94	57	23	15	1	100	100	78	33	11
REGWQ	93	51	21	11	1	100	100	79	33	11
Tukey	92	44	17	6	1	100	100	64	30	11
SMM	86	35	12	5	1	100	100	61	27	11
Sidak	87	32	13	4	1	100	100	61	28	13
Bonferroni	86	32	13	4	2	100	100	61	28	12
Dunnett bilateral	100	70	34	21	3	100	100	81	28	13

QUADRO 3 - Poder dos diferentes testes estatísticos expressos em porcentagem, resultantes da simulação de n=100 e n=33 experimentos com diferenças reais de 30%, 20%, 15%, 10% e 5%, abrangendo 13 tratamentos com r = 5 e r = 15 repetições e coeficientes de variação CV = 15%.

Testes	r = 5, n = 100 experimentos					r = 15, n = 33 experimentos				
	30%	20%	15%	10%	5%	30%	20%	15%	10%	5%
Waller	91	64	44	21	13	100	100	88	64	24
T bilateral	89	60	38	19	10	100	100	82	45	12
Duncan	85	51	31	13	4	100	100	79	33	12
SNK	64	21	9	2	2	100	88	61	24	6
REGWF	56	15	4	1	1	100	85	48	3	3
REGWQ	47	12	2	1	1	100	85	39	3	3
Tukey	40	8	2	0	1	100	58	24	3	0
SMM	37	4	1	0	0	100	52	21	3	0
Sidak	37	4	1	0	0	100	52	21	3	0
Bonferroni	36	4	1	0	0	100	52	21	3	0
Dunnett	77	26	19	5	3	100	94	64	21	3

QUADRO 4 – Poder dos diferentes testes estatísticos expresso em porcentagem, resultantes da simulação de n=100 e n=33 experimentos com diferenças reais de 30%, 20%, 15%, 10% e 5%, abrangendo 13 tratamentos com r = 5 e r = 15 repetições e coeficientes de variação CV= 20%.

Testes	r = 5, n = 100					r = 15, n = 33 experimentos				
	30%	20%	15%	10%	5%	30%	20%	15%	10%	5%
Waller	66	34	20	12	7	100	87	63	30	6
T bilateral	67	33	20	10	4	100	84	57	27	6
Duncan	57	28	14	7	4	100	72	45	21	3
SNK	26	10	5	3	1	90	54	21	15	0
REGWF	14	5	4	0	1	81	30	6	3	0
REGWQ	13	3	2	0	0	75	24	6	3	0
Tukey	21	7	6	2	1	87	42	15	9	0
SMM	15	6	5	0	1	84	39	9	6	0
Sidak	13	3	2	0	0	75	27	6	3	0
Bonferroni	13	2	2	0	0	78	30	6	3	0
Dunnett bilateral	41	14	10	6	2	95	63	24	15	0

QUADRO 5 – Poder discriminativo, em porcentagem, dos vários testes de comparação de médias, com 12 tratamentos, com 3 e 6 repetições e coeficiente de variação CV = 10%.

Testes	Diferenças percentuais em relação ao controle									
	r = 3 (200 experimentos)					r = 6 (100 experimentos)				
	30%	20%	15%	10%	5%	30%	20%	15%	10%	5%
T unilateral	98	75	56	29	16	100	95	79	51	24
T bilateral	96	61	40	20	11	100	93	70	33	17
Duncan	92	54	34	12	8	100	88	67	27	11
SNK	63	23	11	5	2	100	64	33	9	2
Tukey	58	15	6	2	1	100	48	19	4	1
Bonferroni	48	10	4	2	0	100	42	18	3	1
Dunnett bil.	74	32	14	6	3	100	64	37	12	2
Bonferoni M_1	91	59	38	12	5	100	79	48	17	6
Bonferroni M_2	91	63	44	18	11	100	82	55	24	12
Sidak M	84	47	29	10	7	100	78	60	24	13

QUADRO 6 – Poder discriminativo, em porcentagem, dos vários testes de comparação de médias, com 25 tratamentos com 3 e 6 repetições e coeficiente de variação CV = 10%.

Testes	Diferenças percentuais em relação ao controle									
	r = 3 (200 experimentos)					r = 6 (100 experimentos)				
	30%	20%	15%	10%	5%	30%	20%	15%	10%	5%
Waller	97	71	50	27	13	100	91	72	48	22
T unilateral	94	75	57	30	16	100	98	79	50	25
T bilateral	94	68	47	22	13	100	90	69	38	18
Duncan	91	53	33	11	7	100	85	63	26	10
SNK	61	19	6	3	0	99	68	25	5	0
Tukey	44	11	2	1	0	91	41	14	3	0
Bonferroni	35	8	1	1	0	84	37	11	2	0
Dunnett bil.	68	26	12	4	1	100	70	35	8	0
Bonferoni M_1	75	34	15	5	1	100	73	40	12	4
Bonferroni M_2	74	33	15	5	1	99	71	38	13	3
Sidak M	73	34	16	5	1	100	73	40	13	3

QUADRO 7 – Poder discriminativo, em porcentagem, dos vários testes estatísticos de comparação de médias, com 8 tratamentos, com 4 e 8 repetições e coeficientes de variação CV = 10%.

Testes	Diferenças percentuais em relação ao controle									
	r = 4 (400 experimentos)					r = 8 (200 experimentos)				
	30%	20%	15%	10%	5%	30%	20%	15%	10%	5%
Waller	93	68	44	28	9	100	96	79	44	15
T unilateral	98	82	62	38	18	100	97	89	54	21
T bilateral	96	71	48	26	9	100	95	79	44	12
Duncan	94	65	40	20	6	100	94	73	37	11
SNK	80	36	18	6	2	98	79	48	22	4
REGWF	79	34	17	5	1	99	79	47	21	2
REGWQ	76	31	16	4	1	99	76	44	18	1
Tukey	73	28	15	2	1	99	70	37	14	1
Bonferroni	67	21	10	1	0	99	67	36	11	1
Sidak	67	22	11	1	0	99	67	36	12	1
Dunnett unil.	92	56	34	13	4	100	91	64	31	8
Bonferoni M_1	88	54	33	13	4	100	91	64	33	8
Bonferroni M_2	85	50	30	14	4	100	88	64	31	8
Sidak M	92	58	37	19	6	100	93	70	38	13

QUADRO 8 - Eficiência do poder discriminativo em relação ao teste de Student unilateral correspondente a 18 tratamentos, com 3 e 6 repetições e coeficiente de variação CV = 10 %.

Testes	Eficiência em Relação ao Teste de T unilateral									
	r = 3 (200 experimentos)					r = 6 (100 experimentos)				
	30%	20%	15%	10%	5%	30%	20%	15%	10%	5%
T unilateral	1	1	1	1	1	1	1	1	1	1
T bilateral	0,98	0,81	0,71	0,69	0,68	1	0,98	0,89	0,65	0,71
Duncan	0,94	0,72	0,64	0,41	0,5	1	0,93	0,85	0,53	0,46
SNK	0,64	0,31	0,2	0,18	0,12	1	0,67	0,42	0,18	0,14
Tukey	0,59	0,2	0,11	0,07	0,06	1	0,51	0,24	0,08	0,04
Bonferroni	0,49	0,13	0,07	0,07	0	0,99	0,44	0,23	0,06	0,04
Dunnett bil.	0,76	0,43	0,25	0,21	0,19	1	0,67	0,47	0,24	0,08
Bonferoni M_1	0,93	0,79	0,68	0,41	0,31	1	0,83	0,61	0,33	0,25
Bonferroni M_2	0,93	0,84	0,78	0,62	0,69	1	0,86	0,7	0,47	0,5
Sidak M	0,86	0,63	0,52	0,34	0,44	1	0,82	0,76	0,47	0,54

QUADRO 9 - Eficiência do poder discriminativo em relação ao teste de Student unilateral correspondente a 25 tratamentos, com 3 e 6 repetições e coeficiente de variações CV = 10 %.

Testes	Eficiência em Relação ao Teste de T unilateral									
	r = 3 (200 experimentos)					r = 6 (100 experimentos)				
	30%	20%	15%	10%	5%	30%	20%	15%	10%	5%
T unilateral	1	1	1	1	1	1	1	1	1	1
Waller	1,03	0,95	0,9	0,9	0,81	1	0,93	0,91	0,96	0,88
T bilateral	1	0,91	0,82	0,73	0,81	1	0,92	0,87	0,76	0,72
Duncan	0,97	0,71	0,58	0,37	0,44	1	0,87	0,8	0,52	0,40
SNK	0,65	0,25	0,11	0,1	0	0,99	0,69	0,32	0,10	0
Tukey	0,47	0,15	0,04	0,03	0	0,91	0,42	0,18	0,06	0
Bonferroni	0,37	0,11	0,02	0,03	0	0,84	0,38	0,14	0,04	0
Dunnett bil.	0,72	0,35	0,21	0,13	0,06	1	0,71	0,44	0,16	0
Bonferoni M_1	0,8	0,45	0,26	0,17	0,06	1	0,74	0,51	0,24	0,16
Bonferroni M_2	0,79	0,44	0,26	0,17	0,06	0,99	0,72	0,48	0,26	0,12
Sidak M	0,8	0,45	0 28	0,17	0,06	1	0,74	0,51	0,26	0,12

QUADRO 10 - Eficiência dos vários testes estatísticos em relação ao teste de Student unilateral correspondente a 8 tratamentos, com 4 e 8 repetições, e coeficiente de variação CV = 10 %.

Testes	Eficiência em Relação ao Teste de T unilateral									
	r = 4 (400 experimentos)					r = 8 (200 experimentos)				
	30%	20%	15%	10%	5%	30%	20%	15%	10%	5%
Waller	0,95	0,83	0,71	0,74	0,50	0,99	0,89	0,81	0,71	1,00
T unilateral	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
T bilateral	0,98	0,87	0,77	0,68	0,5	0,98	0,89	0,81	0,57	1,00
Duncan	0,96	0,79	0,65	0,53	0,33	0,97	0,82	0,69	0,52	1,00
SNK	0,82	0,44	0,29	0,16	0,11	0,81	0,54	0,41	0,19	0,98
REGWF	0,81	0,41	0,27	0,13	0,06	0,81	0,53	0,39	0,1	0,99
REGWQ	0,78	0,38	0,26	0,11	0,06	0,78	0,49	0,33	0,05	0,99
Tukey	0,74	0,34	0,24	0,05	0,06	0,72	0,42	0,26	0,05	0,99
Bonferroni	0,68	0,26	0,16	0,03	0,00	0,69	0,40	0,2	0,05	0,99
Sidak	0,68	0,27	0,18	0,03	0,00	0,69	0,40	0,22	0,05	0,99
Dunnett unil.	0,94	0,68	0,55	0,34	0,22	0,94	0,72	0,57	0,38	1,00
Bonferoni M_1	0,90	0,66	0,53	0,34	0,22	0,94	0,72	0,61	0,38	1,00
Bonferroni M_2	0,87	0,61	0,48	0,37	0,22	0,91	0,72	0,57	0,38	1,00
Sidak M	0,94	0,71	0,6	0,5	0,33	0,96	0,79	0,7	0,62	1,00