

BONFERRONI'S AND SIDAK'S MODIFIED TESTS

Armando Conagin¹; Décio Barbin^{2*}

¹Instituto Agronômico de Campinas, C.P. 28 - 13020-902 - Campinas, SP - Brasil.

²USP/ESALQ - Depto. de Ciências Exatas, C.P. 9 - 13418-900 - Piracicaba, SP - Brasil.

*Corresponding author <debarbin@esalq.usp.br>

ABSTRACT: Results of practical importance had been discarded testing formulated hypothesis with the aid of statistical analysis of experimental data because of the power of the utilized test. This study compares the power of two Bonferroni's Modified and one Sidak's Modified tests with known tests analyzing 1200 simulated experiments. All differences of means were obtained in relation to the mean of the adopted control to guarantee parametrical magnitude of mean differences. Student's test (type I comparisonwise error) and Waller-Duncan's (Bayesian error) showed the highest percentage of significant differences, followed by Duncan's, BM_2 , S_iM , BM_1 , DunnettU's, S_iN , BN, Dunnettu's, SNK's, REGWF's, REGWQ's, Tukey's, Sidak's and Bonferroni's tests. For differences equal to zero, Student's and Waller-Duncan's test exhibit 5% frequency of rejection of the null hypothesis, in accordance the nominal error I adopted ($\alpha = 0.05$). All other tests had values below 0.05, generally ranging on 0.01 to 0.02 or less. Depending of the number of zero differences and considering the type I experimentwise error I, Student's, Waller-Duncan's and Duncan's tests showed crescent values of errors (> 0.05), proportional to the number of null differences included in the experiment; all other tests exhibit showed of type I experimentwise error < 0.05 , most nearing 0.01-0.02 or less. Efficiency of the three "Modified Tests" was close to DunnettU's test, but higher than the other testes of type I experimentwise error nature (MEER).

Key words: statistical tests, efficiency of tests, multiples comparisons

TESTE MODIFICADO DE BONFERRONI E SIDAK

RESUMO: Para a comprovação de hipóteses, experimentos são conduzidos e os resultados obtidos analisados estatisticamente. Entende-se que, em função dos testes utilizados, muito material de importância prática tem sido descartado. Neste estudo, dois testes de Bonferroni Modificados e um teste de Sidak Modificado foram desenvolvidos, o poder desses testes avaliados através da simulação de 1.200 experimentos e sua eficiência comparada às de testes de significância mais conhecidos. As diferenças das médias foram, todas elas, obtidas em relação à média do tratamento controle de forma a garantir, parametricamente, a magnitude das diferenças. O teste de Student (do tipo erro I, por comparação), Waller-Duncan teste (erro de natureza bayesiana) mostraram mais alta porcentagem de diferenças significativas seguida pelos testes de Duncan, BM_2 , S_iM , BM_1 , Dunnett, SNK, REGWF, REGWQ, Tukey, Sidak e Bonferroni. Para diferenças iguais a zero, os testes de Student, Waller-Duncan exibem frequência de rejeição da hipótese nula próxima a 0,05, de acordo com o erro tipo I adotado ($\alpha = 0,05$). Os outros testes exibem valores $< 0,05$, quase todos com valores entre 0,01 a 0,02 ou menos. Considerando o erro experimental tipo I, por experimento, os testes de Student, Waller-Duncan e Duncan exibem, para diferenças nulas, valores crescentes de erro I, proporcional ao número de diferenças nulas incluídas nos experimentos; os demais testes mostram valores para o erro "experimentwise" $< 0,05$ muitos deles entre 0,01 e 0,02 ou menos. Os três "Testes Modificados" tiveram eficiência próxima do teste de Dunnett unilateral, porém maior eficiência que os demais testes do tipo "experimentwise" (MEER).

Palavras-chave: testes estatísticos, eficiência dos testes, comparações múltiplas

INTRODUCTION

To expand the knowledge in different areas of the experimental sciences, experiments are set up to test hypothesis that best explain the phenomena under investigation. Statistical methods are used to design experiments, to choose group of treatments, to perform analyses of variance, to do statis-

tical tests and to estimate parameters. All those methods provide ways to prove or not the formulated hypothesis.

The tests more frequently used for comparison of treatments (means) are Student's, Duncan's, Student-Newman-Keul's (SNK), Tukey's and Dunnett's. Bonferroni's, Sidak's and Waller-Duncan's testes are used less than often.

Studies have been conducted to evaluate both the power of these tests and the different types of errors involved. The evaluation of the power of each test is obtained through the calculus of the percentage of significative differences in which the probability of error of type I – the rejection of the null hypothesis – adopted is $\alpha = 0.05$ or 0.01 .

More recently, errors type I were classified in type I comparisonwise error and type I experimentwise error, the type I experimentwise error under condition of the general null hypothesis, and the type I experimentwise error under condition of partial hypothesis and maximum experimentwise error rate (MEER) type (SAS, 2004), considering both situations. Both unilateral and bilateral Student's tests and Waller-Duncan's test guarantee protection to comparisonwise errors; SNK's enables protection to general experimentwise error, and Tukey's, Bonferroni's, Sidak's, REGWF's, REGWQ's and Dunnett's answers to errors of MEER type.

Evaluations of the efficiency of various statistical tests were done by Gabriel (1964), Carmer & Swanson (1971; 1973), O'Neill & Wetherill (1971), Boardman & Moffitt (1971), Bernardson (1975), Chew (1977), Petersen (1977), Thomas (1974), Hochberg & Tanhame (1987), and others. The efficiency of different tests more frequently utilized in Brazil was also evaluated by Perecin & Barbosa (1988), Conagin (1998), Conagin et al. (1999), Dos Santos (2000), and Conagin & Pimentel-Gomes (2004). The objective of this paper is evaluating the comparative efficiency of the more frequently utilized tests, and the behavior of the comparisonwise and experimentwise type errors, varying the magnitude of differences, number of treatments, number of replications, and number of degrees of freedom of the residual. The paper also introduces a Bonferroni's Modified and a Sidak's Modified test.

MATERIAL AND METHODS

The power of Student's unilateral (T_1), bilateral (T_2), Duncan's (D), Waller-Duncan's (W), Student-Newman-Keul's (SNK), Tukey's (Tu), Bonferroni's (B), Sidak's (S), Dunnett's (Du) and (DU), Modified Bonferroni's (BM and BM_2), Modified Sidak's (S_M), REGWF and REGWQ, Gin SAS's default method was tested; in the Modified tests, $\alpha = 0.05$. Three groups – G_1 , G_2 and G_3 – included the tests with three, four, six, and eight replications. A discussion on the modifications of the Bonferroni's and Sidak's follow.

Bonferroni's Test

The Bonferroni's test may be used to calculate confidence interval and comparison of means. In this

paper, only differences were studied. Means \bar{x}_i and \bar{x}_j of two treatments with r_i and r_j replications, differ significantly if:

$$|\bar{x}_i - \bar{x}_j| \geq t(\alpha_\beta, df) s \sqrt{\frac{1}{r_i} + \frac{1}{r_j}}$$

where df is the number of degrees of freedom of $QM Res$, $s = \sqrt{QM Res}$, and $t(\alpha_\beta, df)$ is the value of the t-distribution with probability α_β and df degrees of freedom. If $r_i = r_j = r$, then $\frac{1}{r_i} + \frac{1}{r_j} = \frac{2}{r}$.

If a given experiment has t treatments with r replications, and $\alpha = 0.05$ is the global probability for k comparisons, in which $H_0: T_1 = T_2 = \dots = T_t = 0$ is the general null hypothesis, then, for all comparisons between pairs of means, $k = C_t^2 = t(t-1)/2$, and in this case $\alpha_B = \alpha/k$.

If H_0 is true, the probability of not rejecting any difference is $(1 - \alpha_B)^k$. If the global probability adopted is $\alpha = 0.05$ or 0.01 (for t treatments), the probability of rejecting the general H_0 will be:

$$\alpha = 1 - (1 - \alpha_B)^k = 1 - [1 - k\alpha_B + k(k-1)/2 \alpha_B^2 \pm \dots]$$

and considering only the first two terms, $\alpha \approx 1 - [1 - k\alpha_B]$. Then $\alpha_B = \alpha/k$.

The calculus of the significative difference between two means is done applying Student's test with α_B and df degrees of freedom. The test assures $\alpha = 0.05$ for all tests, and the error is of the MEER experimentwise type.

Modified Bonferroni's Test (BM)

Modified Bonferroni's test was proposed by Conagin (1999). The Student's t for the test shall be obtained in Student's Tables, with degree of freedom of the residual and α' probability level, in which $\alpha' = \alpha(1 + P)$, and ordinarily $\alpha = 0.05$ or 0.01 . The calculus of P is presented bellow. In the analysis of variance, the test of H_0 is obtained by the test $F_0 = QMTreat/QMresidual$, in which there are t treatments and r replications (e.g. in randomized blocks design). The critical value of F is F_c , with $(t-1)$ and $(r-1)$ ($t-1$) degrees of freedom, for $\alpha = 0.05$ or 0.01 .

The parametric model is $X_{ij} = M + T_i + B_j + E_{ij}$, in which $i = 1, 2, \dots, t$, and $j = 1, 2, \dots, r$. The expected values are:

$$EQMtreat = \sigma^2 + (r \sum T_i^2)/(t-1)$$

$$EQMresidual = \sigma^2$$

If the general H_0 is true, $T_1 = T_2 = \dots = T_t = 0$. For the sampling model (experimental),

$$QMtreat = s^2 + (r \sum t_1^2)/(t-1)$$

being s^2 the QMresidual. In the Analysis of Variance of the experiment:

$$F_0 = [s^2 + (r \sum t_i^2)/(t-1)]/s^2.$$

Due to the size of experimental error some treatments of real nature are not significative.

If $F_0 > F_c$, H_0 is rejected. Then, there should be one or more treatments $t_e \neq 0$. The parameter of non-centrality of the F distribution, if H_a (alternative hypothesis) is true, is λ (Winer et al., 1991); this value is:

$$\lambda = r \sum T_i^2 / \sigma^2.$$

For a given experiment, $\hat{\lambda} = r \sum t_i^2 / s^2$.

$$F_0 = [s^2 + r \sum t_i^2 / (t-1)] / s^2 = 1 + \hat{\lambda} / (t-1).$$

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

The evaluation of $P(F_0)$ and $P(F_c)$, if H_a is true, may be obtained by PROB F Function of SAS (2004). For fixed t and r:

$$P(F_0) = \int_0^{F_0} F(F_0, (t-1), (t-1)(r-1), \hat{\lambda}) dF$$

$$P(F_c) = \int_0^{F_c} P(F_c, (t-1), (t-1)(r-1), \hat{\lambda}) dF$$

$P = P(F_0) - P(F_c)$ is then defined. This is the probability represented by the area between F_0 and F_c in the F non-central. If F_c is fixed, the area increases if F_0 increases. The area is originated and due to treatments that are far from \bar{X} (general mean) and includes the treatment significant different of \bar{X} .

With α'/k , smaller values than the t value when H_0 is true are obtained. If $t' < t$, the efficiency of the test will be increased.

The corresponding α' for this situation is defined as $\alpha' = \alpha(1+P)$, and then $\alpha_{BM} = \alpha'/k$. In this case, the t value of the Table of Student test for α' value produces $t' < t$, where t' is the value of the Table according to H'_0 (where the new hypothesis have smaller number of treatments in which $t_1 = t_2 = \dots = t_1$, being $1 < t$).

The modified Bonferroni's test uses the t Student test with probability α' , and then $\alpha_{BM} = \alpha'/k$.

Values shown in Tables 1, 2, 5, and 6, use the PROBF Function to calculate P. But since Group G_2 includes 25 treatments, ordinarily $\hat{\lambda} = (F_0 - 1)(t-1)$ produces $\hat{\lambda} > 100$, and then, PROBF Function do not

calculate the $P(F_0)$ and $P(F_c)$. In this case, calculated Tables (Conagin, 2001) for $\alpha = 0.05$, t, r and F_0/F_c included in the cells the corresponding P value are used. If $\hat{\lambda} > 100$ and $F_0/F_c > 7$, an proximate value for P using $F_0/F_c = 7$ and $\hat{\lambda} = 100$ is used. The justificative is that if $F_0/F_c > 7$ the area (and then P) should be greater than $F_0/F_c = 7$, and then adopted P is a conservative value.

Second Modified Bonferroni's Test (BM₂)

If in the Analysis of variance $F_0 > F_c$ (general H_0 rejected), there should be a treatment parametrically different of zero. It is possible to evaluate a by \hat{a} and use it to calculate BM₂, as follows.

If $F_0 > F_c$ the ANOVA is performed and the significant differences between two means by the Student's test (comparisonwise type I error) are calculated. The \hat{a} number of significant differences is used as an estimate of the true number of parametrical differences between treatments. This happens because, in general, experiments are performed to evaluate responses of new "treatments", supposedly superior to treatments ordinarily used (especially in agronomy, animal science and veterinary medicine, medical, biological, and industrial research, and some other areas in which a new or new treatments are supposed to be better, in some way, than currently used treatments). The \hat{a} number of significant differences should be $\hat{a} < a$. Due to the size of experimental error, some treatments actually different from zero but of small values, result in non-significant differences, and then \hat{a} is, probably, a conservative estimate of the true a value ($\hat{a} < a$).

The probability modified Bonferroni's (BM₂) is then:

$$\alpha_{BM_2} = \alpha / (k - \hat{a}).$$

The modified Bonferroni's test behaves similarly to tests of the type MEER, as can be possibly seen in the columns 0% (exp) of Tables 1, 2, 3, 4, 5, and 6. Their frequencies are very much alike the corresponding Dunnett's frequency and it is well known that Dunnett's test is of MEER type.

Sidak's Test (S_i)

The Sidak's test uses a probability α_s and, similarly to Bonferroni's, use the Student t test with level α_s and df degrees of freedom of the residual. The α_s value is calculated from:

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

in which α_s is the global probability and k is the number of comparisons between means.

Table 1 - Power of different statistical tests for Group 1, resultant of simulation of 200 experiments containing twelve treatments, three replications, CV = 10%, for differences of 40%, 35%, 30%, 25%, 20%, 15%, 10%, 5% and 0%, in comparison with a control.

Test										(cp)	(exp)
	40	35	30	25	20	15	10	5	0	0	
----- % -----											
T ₂	100.0	98.0	96.0	82.5	61.5	40.5	20.0	11.0	2.5	2.5	
D	99.5	93.5	92.0	73.0	54.5	34.0	12.0	8.5	1.5	1.5	
SNK	93.0	84.0	63.5	41.5	23.5	11.0	5.0	2.0	1.0	1.0	
Tu	87.0	71.0	58.5	30.0	15.5	6.5	1.5	1.5	0.0	0.0	
B	80.0	59.5	48.5	26.0	10.5	4.5	1.5	0.5	0.0	0.0	
Du	98.0	87.0	74.0	51.5	32.0	14.0	6.0	2.5	0.5	0.5	
BN	97.0	90.0	78.0	53.5	35.0	33.0	10.0	2.0	0.5	0.5	
BM	99.0	94.5	91.5	77.0	59.5	38.0	12.5	5.5	0.5	0.5	
BM ₂	98.5	94.0	91.0	77.5	63.0	44.0	18.5	11.0	1.5	1.5	
SiN	97.5	90.5	79.0	56.0	36.5	33.0	7.0	3.0	1.0	1.0	
SiM	97.5	91.0	84.5	64.5	46.5	28.5	10.0	7.0	1.5	1.5	

Table 2 - Power of different statistical tests for Group 1, resultant of simulation of 100 experiments containing twelve treatments, six replications, CV = 10%, for differences of 40%, 35%, 30%, 25%, 20%, 15%, 10%, 5% and 0%, in comparison with a control.

Test										(cp)	(exp)
	40	35	30	25	20	15	10	5	0	0	
----- % -----											
T ₂	100.0	100.0	100.0	100.0	93.0	70.0	33.0	17.0	3.0	3.0	
D	100.0	100.0	100.0	97.0	88.0	67.0	27.0	11.0	0.0	0.0	
SNK	100.0	100.0	98.0	90.0	64.0	33.0	9.0	2.0	0.0	0.0	
Tu	100.0	100.0	97.0	77.0	48.0	19.0	4.0	1.0	0.0	0.0	
B	100.0	99.0	91.0	76.0	42.0	18.0	3.0	1.0	0.0	0.0	
Du	100.0	100.0	100.0	90.0	64.0	37.0	12.0	2.0	0.0	0.0	
BN	100.0	100.0	99.0	94.0	72.0	43.0	14.0	2.0	0.0	0.0	
BM	100.0	100.0	100.0	96.0	79.0	48.0	17.0	6.0	0.0	0.0	
BM ₂	100.0	100.0	100.0	96.0	82.0	55.0	24.0	12.0	2.0	2.0	
SiN	100.0	100.0	100.0	95.0	74.0	45.0	16.0	3.0	0.0	0.0	
SiM	100.0	100.0	100.0	96.0	78.0	60.0	24.0	13.0	2.0	2.0	

Ordinarily, if the interest lies in all comparisons between two means $k = C_t^2 = t(t-1)/2$. To compare each treatment with a control, $k = t - 1$; this type appear as S_iN in the tables. The value α_s is greater than α_B and then the corresponding t_s is smaller than t_B . So Sidak's test is a little more efficient than Bonferroni's. For instance, for $\alpha = 0.05$, $k = 10$, $\alpha_s = 0.0512$, and $\alpha_B = 0.05$. Therefore, t_s is smaller than t_B .

Modified Sidak's Test (S_iM)

Similarly to Second Bonferroni's Modified (BM₂), after performing an ANOVA of a given experi-

ment, if $F_0 > F_C$ (H_0 general hypothesis rejected), significant differences are calculated by the t test; the number of significant differences \hat{a} is an estimate of the number of all the actual differences a . Then:

$$\alpha_{S_iM} = 1 - (1 - \alpha)^{1/(k-\hat{a})}$$

This test tends to be a little more effective than the correspondent Bonferroni's (BM₂), because $\alpha_{S_iM} > \alpha_{BM_2}$. The calculus of significative differences in relation to a control uses t test with $k=(t-1)-\hat{a}$. If interest lies in all differences between two means, $k=(t(t-1)/2)-\hat{a}$.

Table 3 - Power of different statistical tests for Group 2, resultant of simulation of 200 experiments containing 25 treatments, three replications, CV = 10%, for differences of 40%, 35%, 30%, 25%, 20%, 15%, 10%, 5% and 0%, in comparison with a control.

Test	(cp)									(exp)
	40	35	30	25	20	15	10	5	0	0
----- % -----										
W	99.5	98.5	97.0	91.0	71.5	50.0	27.0	13.5	6.5	33.0
T ₂	99.5	97.0	94.0	87.0	68.0	47.5	22.0	13.5	5.3	23.5
D	99.5	97.0	91.0	79.0	53.5	33.5	11.5	7.0	2.5	12.5
SNK	95.0	79.0	61.0	38.0	19.5	6.0	2.5	0.0	0.8	1.5
Tu	80.0	59.0	44.0	26.0	11.0	1.5	1.0	0.0	0.0	0.5
B	76.0	58.5	35.5	21.0	8.5	1.0	1.0	0.0	0.0	0.0
Du	97.5	84.0	68.0	49.5	26.5	12.5	4.0	0.5	0.8	2.5
BN	96.0	84.5	69.0	50.5	29.0	11.5	3.0	0.5	0.5	1.0
BM	96.5	87.0	75.0	56.5	34.5	15.0	5.5	0.5	0.5	1.5
BM ₂	97.0	85.5	74.5	54.5	33.5	14.5	5.0	1.0	0.5	1.0
SiN	98.0	86.5	71.5	51.5	29.5	12.5	3.5	0.5	0.1	1.0
SiM	98.0	89.5	73.5	54.5	34.5	15.5	5.5	1.0	0.5	2.0

Table 4 - Power of different statistical tests for Group 2, resultant of simulation of 100 experiments containing 25 treatments, six replications, CV = 10%, for differences of 40%, 35%, 30%, 25%, 20%, 15%, 10%, 5% and 0%, in comparison with a control.

Test	(cp)									(exp)
	40	35	30	25	20	15	10	5	0	0
----- % -----										
W	100.0	100.0	100.0	97.0	91.0	72.0	48.0	22.0	8.6	36.0
T ₂	100.0	100.0	100.0	100.0	90.0	69.0	38.0	18.0	6.5	25.0
D	100.0	100.0	100.0	98.0	85.0	63.0	26.0	10.0	3.0	11.0
SNK	100.0	100.0	99.0	92.0	68.0	25.0	5.0	0.0	0.5	1.0
Tu	100.0	99.0	91.0	68.0	41.0	14.0	3.0	0.0	0.5	1.0
B	100.0	97.0	84.0	66.0	37.0	11.0	2.0	0.0	0.0	0.0
Du	100.0	100.0	100.0	91.0	70.0	35.0	8.0	0.0	0.5	4.0
BN	100.0	99.0	99.0	91.0	71.0	37.0	9.0	0.0	0.1	0.1
BM	100.0	100.0	100.0	95.0	73.0	40.0	12.0	4.0	0.3	4.0
BM ₂	100.0	99.0	99.0	92.0	71.0	38.0	13.0	3.0	0.4	2.0
SiN	100.0	100.0	100.0	91.0	71.0	36.0	9.0	0.0	0.0	1.0
SiM	100.0	100.0	100.0	94.0	73.0	40.0	13.0	3.0	1.0	1.0

The behavior of the application of Modified Sidak's test can be evaluated in Tables 1, 2, 3, 4, 5, and 6.

Power of the different tests in the groups G₁, G₂ and G₃

To evaluate the power of Student's t test T₂ (bilateral), Duncan's, SNK's, Tukey's, Bonferroni's, Dunnett's, BN's, BM's, BM₂'s, S_iN's and S_iM's, simulations of two hundred experiments, r = 3 replications and considering differences of 0.4; 0.3; 0.25; 0.15; 0.1; 0.05 and 0.0 (both cp and exp); and one hundred ex-

periments, r = 6 replications, groups G₁ (12 treatments) and G₂ (25 treatments), were analyzed. For both groups, CV = 10%. In group G₂, a Waller-Duncan's test (W) was included. The "default criterion" (SAS, 2004) was used for the ANOVA and for the Modified Bonferroni's and Sidak's tests, calculus of significant differences ($\alpha = 0.05$) were done separately by the authors.

In group G₃ (Tables 5 and 6), parametrical differences were 0.3, 0.2, 0.15, 0.1, 0.05, and 0.0 (both cp and exp), for r = 4 (400 experiments) and r = 8 (200 experiments). Once again "default criterion" was used

Table 5 - Power of different statistical tests for Group 3, resultant of simulation of 400 experiments containing 8 treatments, 4 replications, CV = 10%, for differences of 30%, 20%, 15%, 10%, 5% and 0%, in comparison with a control.

Test	(cp) (exp)						
	30	20	15	10	5	0	0
----- % -----							
W	93.3	68.0	44.0	27.8	9.3	4.4	8.5
T ₁	98.0	82.5	62.0	38.3	17.5	5.3	12.7
T ₂	96.0	71.5	48.3	26.3	9.0	4.6	9.3
D	94.5	65.0	39.8	20.0	5.5	4.3	9.0
SNK	79.8	36.0	18.3	5.8	1.8	1.6	3.0
REGWF	79.0	34.3	17.5	5.3	0.8	0.8	1.5
REGWQ	76.5	30.8	16.3	3.8	1.3	0.6	1.5
Tu	72.8	28.3	15.3	2.3	0.5	0.3	0.8
B	67.0	21.3	10.0	1.3	0.3	0.1	0.5
Si	67.5	21.8	10.5	1.3	0.3	0.1	0.2
DuU	91.8	56.3	33.5	13.3	3.8	2.3	4.3
BN	87.3	48.3	29.5	11.8	2.5	1.5	4.3
BM	87.8	53.5	32.5	13.0	3.8	2.4	4.3
BM ₂	85.3	50.0	30.3	13.5	3.5	1.8	3.3
SiN	88.0	49.8	27.5	10.0	2.5	1.2	1.8
SiM	92.0	57.8	37.5	19.0	6.0	2.5	4.8

for the ANOVA tests and $\alpha = 0.05$ for the modified Bonferroni's and Sidak's test. Data this group were also submitted to REGWF's, REGWQ's, Student's unilateral (T₁), Dunnett's unilateral (DU) tests, and Sidak's basic (S_i) test, in which $k = t(t-1)/2$.

For large differences (i.e. 0.4, 0.35, 0.3, and 0.25) on G₁ and G₂ (Tables 1 and 2), and G₂ alone (Tables 3 and 4), tests showed high power, with little differences among them. Differences in the power of the various tests increased for differences 0.2, 0.15, and 0.1 to values often and more ordinarily obtained in the research data analysis. In groups G₁ and G₂, power of tests were, in order, Waller-Duncan (W), Student bilateral (T₂), Duncan (D), SNK, S_iM, BM, BM₂, S_iN, BN, Dunnett, Tukey and Bonferrooni.

Regarding error of type I comparisonwise [column 0% (cp)], Student's and Waller-Duncan's tests showed values $\alpha \approx 0.05$, the nominal error adopted. The other tests showed $\alpha < 0.05$, most of them around 0.01 or smaller. The error I per experiment (column 0% exp; Tables 3 and 4) of Student's (T₂), Waller-Duncan's and Duncan's tests were much higher than 0.05; other tests showed errors smaller than 0.05, most nearing 0.01, or smaller.

Table 6 - Power of different statistical tests for Group 3, resultant of simulation of 200 experiments containing twelve treatments, 8 replications, CV = 10%, for differences of 30%, 20%, 15%, 10%, 5% and 0%, in comparison with a control.

Test	(cp) (exp)						
	30	20	15	10	5	0	0
----- % -----							
W	100.0	96.5	79.5	44.5	15.5	5.8	8.0
T ₁	100.0	97.5	89.5	54.5	21.0	5.5	9.0
T ₂	100.0	95.5	79.0	44.0	12.0	4.8	7.5
D	100.0	94.5	73.5	37.5	11.0	3.8	5.5
SNK	99.5	79.5	48.5	22.0	3.5	1.5	3.0
REGWF	99.5	79.0	47.0	21.5	2.0	1.3	2.5
REGWQ	99.5	75.5	44.0	18.0	0.5	1.0	2.0
Tu	99.5	70.0	36.5	14.0	1.5	0.8	0.5
B	99.5	67.5	35.5	11.0	1.0	0.2	0.0
Si	99.5	67.5	35.5	12.0	1.0	0.3	0.5
DuU	100.0	91.5	63.5	31.5	8.0	1.3	2.5
BN	100.0	85.0	48.5	26.5	5.5	0.8	2.0
BM	100.0	91.5	64.0	33.5	8.0	1.3	2.0
BM ₂	100.0	88.0	64.5	31.0	8.5	1.0	1.5
SiN	99.5	87.0	61.5	27.5	6.0	1.0	1.5
SiM	100.0	93.0	70.0	38.0	13.5	3.0	0.4

In the group G₃ (t = 8; n = 4 and 8; CV = 10%), there were differences of efficiency among tests for values of 0.3 or less (Table 5). In the range of 0.1-0.2, Student's unilateral (T₁) was slightly superiority to Waller-Duncan's (W) and Student's bilateral (T₂), that increases when r = 8 (Table 6). The overall order of efficiency registered was: T₁, T₂, W, D, S_iM, DU, BM, BM₂, S_iN, BN, REGWF, REGWQ, Tu, Si and B.

Regarding errors of type I comparisonwise [column 0% (cp)], W, T₁, T₂, and D, showed $\alpha = 0.05$, the nominal error adopted. The other tests showed $\alpha < 0.05$, most nearing 0.01. For errors of type I experimentwise [column 0% (exp)], T₁, T₂, W, and D, showed α values well above 0.05; the other tests showed α values smaller than 0.05, most nearing 0.01 or less. The power of all the tests increased when the number of replications increased. It is therefore easier to compare values in the Tables 1, 3, and 5, with the corresponding values in Tables 2, 4, and 6, respectively.

REFERENCES

- BERNARDSON, C.S. Type I error rates when multiple comparison procedures follows a significant F test of Anova. *Biometrics*, v.31, p.337-340, 1975.

- BOARDMAN, T.J.; MOFFITT, D.R. Graphical Monte Carlo type I error rates for multiple comparison procedures. **Biometrics**, v.27, p.738-744, 1971.
- CARMER, S.G.; SWANSON, M.R. Detection of differences between means, a Monte Carlo study of five multiple comparison procedure. **Agronomy Journal**, v.63, p.940-945, 1971.
- CARMER, S.G.; SWANSON, M.R. Evaluation of ten multiple comparison procedures by Monte Carlo methods. **Journal of American Statistical Association**, v.68, p.66-74, 1973.
- CHEW, V. **Comparisons among treatment means in an analysis of variance**. Beltsville: USDA, 1977. 69p. (Bulletin, H6).
- CONAGIN, A. Discriminative power of Modified Bonferroni's Test. **Revista de Agricultura**, v.73, p.31-46, 1998.
- CONAGIN, A. Discriminative power of the Modified Bonferroni's Test under general and partial null hypothesis. **Revista de Agricultura**, v.74, p.117-126, 1999.
- CONAGIN, A. Tables for the calculation of the probability to be used in the Modified Bonferroni's Test. **Revista de Agricultura**, v.76, p.71-83, 2001.
- CONAGIN, A.; PIMENTEL-GOMES, F. Escolha adequada dos testes estatísticos para comparações múltiplas. **Revista de Agricultura**, v.79, p.288-295, 2004.
- CONAGIN, A.; IGUE, T.; NAGAI, V. **Poder discriminativo de diferentes testes de médias**. Campinas: Instituto Agronômico, 1999. (Boletim Científico, 44).
- DOS SANTOS, C. Novas alternativas de testes de agrupamento avaliadas por meio de simulação Monte Carlo. Lavras: UFLA, 2000. 85p. (Dissertação – M.S.).
- 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.
- HOCHEBERG, Y.; TAMHAME, A.C. **Multiple comparisons procedures**. New York: John Wiley & Sons, 1987. 450p.
- O'NEILL, R.; WETERILL, G.B. The present state of multiple comparison methods. **Journal of the Royal Statistical Society**, v.33, p.218-250, 1971.
- PERECIN, D.; BARBOSA, J.C. Uma avaliação de seis procedimentos para comparações múltiplas. **Revista de Matemática e Estatística**, v.6, p.95-103, 1988.
- PETERSEN, R.G. Use and misuse of multiple comparison procedures. **Agronomy Journal**, v.69, p.205-208, 1977.
- SAS INSTITUTE. **SAS/STAT 9.1: User's Guide**. Cary: SAS Institute Inc., 2004. 5136p.
- THOMAS, D.H. Error rates in multiple comparisons among means results of a simulation exercise. **Applied Statistics**, v.23, p.284-294, 1974.
- WALLER, R.A.; DUNCAN, D.B. A Bayes rule for the symmetric comparison problem. **Journal of the American Statistical Association**, v.67, p.253-255, 1972.
- WINER, B.J.; BROWN, D.R.; MICHELS, K.M. **Statistical principles in experimental designs**. 3.ed. New York: McGraw-Hill, 1991. 1057p.

Received August 03, 2005

Accepted December 28, 2005