Revista de Agricultura

Vol. 74 Setembro/1999 Nº 2

DISCRIMINATIVE POWER OF THE MODIFIED BONFERRONI'S TEST UNDER GENERAL AND PARTIAL NULL HYPOTHESIS

Armando Conagin[1]

INTRODUCTION

The problem of choosing the most adequate statistical test for comparisons of means of two treatments in research work is dependent on the type of error of first species adopted, if **comparisonwise** or **experimentwise**, this one under general and/ or partial null hypothesis.

It is well known that LSD's and Duncan's multiple range test are of the **comparisonwise** type, SNK's is of the **experimentwise** type under general null hypothesis and that Tukey's, Bonferroni's and Dunnett's test

1 Agronomist and Applied Statistician, retired from the Instituto Agronômico, Campinas, SP, Brazil.

are of the **experimenwise** type under general and / or partial null hypothesis (SAS, 1990).

We are designating Fisher's Test by (LSD), Duncans' by (D), Student-Newman-Keuls by (SNK),Tukey's by (Tu), Bonferroni's by (B), Dunnett's by (Du), the Modified Bonferroni's under general null hypothesis by (BM) and for partial null hypothesis by (BMP). A good discussion on the use of different tests is presented in CHEW (1977), STEEL & TORRIE (1981), in the SAS (1990) and in WINNER et al. (1991).

CARNER & SWANSON (1973) and PERECIN & BARBOSA (1988) performed simulations on various of these tests and pointed out conclusions that agree with those here reached.

In a previous paper (CONAGIN, 1998) the same tests now discussed were studied, but with the Bonferroni's Modified Test only under the general null hypothesis. This new paper adds to the precedent work the consideration of Bonferroni's Modified Test under a partial null hypothesis.

MATERIAL AND METHODS

The theorethical model used, of the randomized block type, was:

$$Y_{ij} = M + T_i + B_j + E_{ij} \qquad i = 1,2,.....t \; ; \qquad j = 1,2,.....r.$$

The values of the parameters were: M = 3000, the T_1's values were 40%, 30%, 20%, 10% etc, of the M value. Three groups, G_1, G_2 and G_3 were studied. For G_1, T_1 = 40%, T_2 = 30% , T_3 = 20%, T_4 = 10%, T_5 = 0% and T_6 =0% (control). In this group K=4, t=6, and all comparisons with the control were performed, for CV=10%, r=3 and r=6.

For G_2, T_1 = 40%, T_2 = 30% , T_3 = 20%, T_4 = 10%, T_5 = 0% , T_6 =0% (control), T_7 = 5%, T_8 = 3%, T_9 = 1%, T_{10} = -5%, T_{11} = -3% and T_{12} = -1%. Here, again, K=4, t=12, CV=10% and r were r=3 and r=6 .

For G_3, $T_1 = 40\%$, $T_2 = 30\%$, $T_3 = 20\%$, $T_4 = 10\%$, $T_5 = 0\%$, $T_6 = 0\%$ with other Ti values totaling t=18; again, CV=10% and r=3 and r=6.

The Bj's values were the same in all sets.

The Eij's were obtained throughout the "RANNOR FUNCTIONS" of the SAS, 1990.

We performed 200 experiments for each group when r=3 and 100 experiments for each group when r=6, joining 1+2, 3+4, etc. single experiments.

On the whole 600 experiments were obtained for r=3 type and 300 experiments for r=6 type. We used SAS Program (SAS, 1990) to perform the simulations, the analysis of variance and the tests already specified.

The difference of treatments with the control 6 was: 40% with treatment 1, 30% with 2, 20% with 3 and 10% with 4. The significance of these differences can evaluate the discriminative power of tests performed.

THE MODIFIED BONFERRONI TEST

The usual Bonferroni Test may be used to stablish confidence intervals or as a statistical test of significance for K comparisons chosen *a priori*.

The means of two treatments, $\bar{y}_1$ and $\bar{y}_2$ are considered different according to Bonferroni's test if

$$|\bar{y}_1 - \bar{y}_2| \geq t(\delta_B, df)s\sqrt{2/r} \qquad \text{(SAS, 1990),}$$

Using the Student **t** distribution with δ_B level of significance, **df** degrees of freedom; **s** is the standard deviation, and **r** the number of replications for each mean.

To calculate δ_B, if a is the joint level of significance of the K comparisons (commonly 0,05 or 0,01), then, for each of the K stipulated comparisons, $\delta_B = \alpha/K$.

If the number of treatments in the experiments is **t** , the δ_B value adequate for comparisons of all possible differences between two means should be $\delta_B = \alpha/t(t-1) \div 2$, being K, in this case the number of all possible combinations of the **t** treatments taken two at a time.

In The Modified Bonferroni Test, under general null hypothesis for the comparisons of K differences between two means, chosen *a priori* we should start testing the general null hypothesis through the Analysis of Variance. If the null hypothesis is rejected at α level, that is, if F_O obtained is greater than the critical level F_C, we perform the Modified Bonferroni Test (general).

The test assumes that

$$\delta_{BM} = \alpha\,(1 + P\,(F))\,/\,K \qquad \text{(A)}$$

The calculation of P (F) is obtained as follows:

Let us take F_o = MS treat. / MS residual in the analysis of variance of the experiment. We know that the populational expectations for these mean squares, under fixed model, are:

$$\text{EMS treats} = \sigma^2 + (\,r\,\Sigma\,T_i^{\,2})\,/\,(\,t-1\,);$$

$$\text{EMS residual} = \sigma^2.$$

If H_O (general null hypothesis) is true, then, for the population ,

$T_1 = T_2 = \ldots = T_t = 0$ and for the experiment:

$t_1 = t_2 = \ldots = t_t = 0$ ..

If H_a (general alternative hypothesis) is true: at least some of the Ti's are different of zero.

The non-centrality parameter of the non-central F distribuition is $\lambda = r \Sigma Ti^2 / \sigma^2$ (WINNER et al., 1991).

In the analysis of the experiment (sample),

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

An estimate $\hat{\lambda}$ of λ is

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

If the critical value of F for the rejection of H_O is the value F_C and $F_O > F_C$, than H_O is rejected.

The probability of values smaller than F_O or F_C if H_a is true may be calculated by PROBF FUNCTION of the SAS Program (SAS-1990).

We define $P(F) = P(F_0) - P(F_C)$; it represents the probability of the values of F smaller than F_0 and greater than F_C if H_a is true. The interval of variation of P(F) is $0 < P(F) < 1$.

As an example, suppose that in an experiment, r = 3, df = 10, t = 6, $F_0 = 9.72$, $F_C = 3.33$ and K = 4. Since $F_0 > F_C$, the null hypothesis shall be rejected. The calculations needed are:

$$\hat{\lambda} = (t-1)(F_o - 1) = 5(9.72 - 1) = 43.60,$$

$$P(F_0) = \int^{F_0} g(F_o = 9.72; t-1 = 5; (t-1)(r-1) = 10; \hat{\lambda} = 43.60) dx = 0.4682$$

$$P(F_c) = \int_0^{F_c} g(F_o = 9.72; t-1 = 5; (t-1)(r-1) = 10; \hat{\lambda} = 43.60) dx = 0.0155$$

Then:

$P(F) = P(F_0) - P(F_C) = 0.4683 - 0.0155 = 0.4527.$

Using formula (A)

$\delta_{BM} = 0.05(1 + 0.4527)/4 = 0.01816.$

For the Bonferroni test in the same conditions it would be:

$\delta_B = 0.05 / 4 = 0.01250.$

The interval of variation of δ_{BM} is: $\delta_B < \delta_{BM} < 2\delta_B$.

Normally the PROBF FUNCTION of the SAS SYSTEM calculates $P(F_0, (t-1), df, \hat{\lambda})$ for values of $\hat{\lambda}$ smaller than 100. In cases in which the $\hat{\lambda}$ value obtained is greater than 100, an approximate solution for values in the range $100 < \hat{\lambda} < 120$, for example , is obtained when we substitute the actual $\hat{\lambda}$ value by 99.99, and calculate a new value $F'_0 = (F_0 \times 99.99) / \hat{\lambda}$.

The $P(F_0', (t-1), df, 99.99)$ obtained is a conservative estimate of P, that may be used until the range of calculated values is enlarged.

The argument to use δ_{BM} is based on the following reasoning: "If the F ratio MS treat / MS residual is large ($F_0 > F_C$), the null hypothesis is rejected and there is "evidence" of the existence of differences among treatments; therefore we do not need to be so rigorous in utilizing $\alpha = 0.05$ for the joint comparisons because we already accepted that H_0 is false. We consider P(F) as the weight of "evidence" that H_a is true. So we include P(F) to liberate the joint probability and use $\alpha(1 + P(F))$ as the joint level for K *a priori* comparisons. So for each comparison $\delta_{BM} = \alpha(1 + P(F)) / K$.

To be consistent, we should calculate the General Modified Bonferroni Test (BM) only if H_0 is rejected (H_0 is false). The structure of the test proposed incorporates the "weight of evidence" of the veracity of H_a through the introduction of the F_0 value obtained in the analysis of variance for the calculation of the t value, used to judge chosen *a priori* K comparisons between two means. The results presented in Tables 1 and 2 regarding BM were obtained in this way.

For the Partial Modified Bonferroni Test (PMB) we calculate F_0 including in the MS treat., only the K + 1 treatments involved in the test to perform the K comparisons, and use this value to calculate $\hat{\lambda}$, P(F) and δ_{BM}.

Table 1. Results obtained when r = 3, CV = 10% for K = 4, for groups G 1, G 2 and G 3, and for df = 10(t = 6), df = 22 (t = 12) and df = 34 (t = 18), for differences of 40%, 30%, 20%, 10%, of the respective treatment mean and the control mean. Values in the cells represent the "discriminative power" in percentage in 200 simulated experiments.

	40%			30%			20%			10%		
df	10	22	34	10	22	34	10	22	34	10	22	34
LSD	99	99	100	94	92	96	56	60.5	70	19	22.5	23.5
D	99	97.5	99.5	91.5	88	95	51	49.5	59	18.5	16.5	13
SNK	93.5	92	95	73	63	65.5	30.5	26	19.5	11.5	3.5	2
Tu	91	87.5	89.5	62.5	52	53	17	16.5	13.5	2	2.5	1.5
B	80.5	87	82.5	50	44.5	45.5	10.5	14	10.5	0.5	2	1
Du	95.5	94	97	71.5	73.5	74	26	33	33.5	8.5	6	4.5
BM	97.5	97	99.5	78	84.5	90	33	45	56.5	9	10	7.5
PMB	95	95.5	98.5	74.5	83	88	32	44.5	55.5	9	11	11

RESULTS

Tables 1 and **2** present the results of the comparison of the different tests, including PMB.

In those Tables the cells represent the percentage of significant results obtained in the comparison between the specified mean with the control, in a large group of experiments. PMB presented always results very close to the ones given by BM test. We believe that if the K + 1 treatments chosen *a priori* presented values smaller than a lot of other treatments in the experiment, the PMB test would not agree so closely with BM, because F_0 and $\hat{\lambda}$ should be different of the values obtained in the BM test.

For the present situation, based on results of **Table 1** and **Table 2** when K = 4 and for different numbers of treatments, replications and degrees of freedom of the residual, we may assume that both types of the Modified Bonferroni's test behaved better than the group of "experimentwise tests" considered , because the new tests improved the "discriminative power" with higher efficiency in the detection of significance of differences studied.

RESUMO

O poder discriminativo de um novo teste estatístico, em condições de hipótese de nulidade geral ou parcial, designado por **Teste de Bonferroni Modificado**, foi comparado com outros testes muito utilizados na comparação de médias.

Mostra-se que, para comparação de K diferenças escolhidas *a priori,* o novo teste é mais discriminativo que os testes de Student – Newman-Keuls (SNK), de Tukey (Tu), de Bonferroni (B) e de Dunnett (Du), todos eles do tipo de erro **por experimento**, sendo menos discriminativo que os testes de Fisher (LSD) e o de comparação múltipla de Duncan (D) que são do tipo de erro **por comparação**.

Table 2. Results obtained when r = 6, CV = 10% for K = 4, for groups G 1, G 2 and G 3, and for df = 20(t = 6), df = 44 (t = 12) and df = 68 (t = 18), for differences of 40%, 30%, 20%, 10%, between the respective treatment mean and the control mean. Values in the cells represent the "discriminative power" in percentage in 100 experiments each one resultant of grouping two single experiments.

	40%			30%			20%			10%		
df	20	44	68	20	44	68	20	44	68	20	44	68
LSD	100	100	100	100	100	99	90	92	95	35	41	40
D	100	100	100	99	99	99	87	87	90	35	36	25
SNK	100	100	100	98	99	99	72	67	68	26	10	8
Tu	100	99	100	98	90	96	56	48	50	7	7	6
B	99	99	100	97	89	95	47	39	41	4	5	4
Du	100	100	100	98	99	99	72	66	76	12	20	13
BM	100	100	100	98	99	99	77	81	87	17	35	26
PMB	100	100	100	99	99	99	76	80	88	17	35	25

Os resultados do poder discrimanativo dos testes estatísticos comparados para diferentes números de tratamentos, com K=4, para coeficiente de variação (CV) de 10% e número de repetições (r) três e seis, constam das **Tabelas 1 e 2**.

Palavras-chave: Poder discriminativo de testes, testes estatísticos usados no melhoramento vegetal.

SUMMARY

The discriminative power of a new statistical test, called **Modified Bonferroni's Test**, under general or partial null hypothesis, was compared with several of the most utilized statistical tests adequate for pairwise comparisons.

It is shown that for K=4 comparisons chosen "a priori", CV= 10% and different number of treatments and replications, the new test was more discriminative than Student-Newman-Keuls, Tukey's, Bonferroni's and Dunnett's, all of **experimentwise** type of error, loosing to Fisher's LSD test and Duncan's multiple range test which are of **comparisonwise** type.

Key words: Discriminative power of tests, comparison of tests used in plant breeding.

REFERENCES

CARNER, S.G. & M.R. SWANSON, 1973. Evaluation of Ten Pairwise Multiple Comparison Procedures by Monte Carlo Methods. **Jour. Amer. Statis. Assoc., 68:** 66-74.

CHEW, V., 1977. **Comparison Among Treatment Means in an Analysis of Variance**. Washngton, USDA, Agricultural Research Service. 60p.

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

PERECIN, D. & J.C. BARBOSA, 1988. Uma Avaliação de Seis Procedimentos para Comparações Múltiplas . **Rev. Mat. Estat.**, São Paulo, **6:** 95-103.

SAS, 1990. **SAS User's Guide.** Carg. SAS Institute. cap. 11, 12 e 24.

STEEL, R.G.D. & J.H. TORRIE, 1981. **Principles and Procedures of Statistics**. McGraw-Hill. 663p.

WINNER, B.J.; D.R. BROWN & K.M. MICHELS, 1991. **Statistical Principles in Experimental Designs**. McGraw-Hill. 1057p.