A MULTIPLE COMPARISON SIGN TEST: TREATMENTS VERSUS CONTROL

ROBERT G. D. STEEL
Mathematics Research Center, U. S. Army
University of Wisconsin

Let $(X_{oj}, X_{1j}, \cdots, X_{kj})$ be the result of a single trial, where the subscript o is associated with a control and the subscripts $1, \cdots, k$ with treatments. To test the joint hypothesis $P(X_{ij}-X_{oj}>0)=1/2 =P(X_{ij}-X_{oj}<0)$, all i, compute the test criterion $(r_1, \cdots, r_k)$ where r_i is the number of times $X_{ij}-X_{oj}$ is negative in n trials. A method for computing the distribution of $(r_1, \cdots, r_k)$ is illustrated. Exact probability distributions of min r_i are given for $k=2$, $n=4(1)10$ and $k=3$, $n=4(1)7$. It is conjectured that $2(\min r-n/2)/\sqrt{n}$ is distributed approximately as Dunnett's t. Tables based on this conjecture are computed and values are seen to agree well with comparable values from the exact distribution.

1. INTRODUCTION

The analysis of variance is an important tool in the analysis of data. A significant F is evidence to infer real treatment differences but gives no information on their location. The need to locate real differences first resulted in independent comparison procedures of which the ultimate calls for independent single degree of freedom comparisons. However, in practice, the most meaningful set of comparisons may not be an independent set. This need gave rise to a number of multiple comparison procedures for non-independent comparisons. Among such multiple comparison procedures is that of Dunnett [2] for comparing several treatments with a control. Dunnett also provides a method for computing a joint set of confidence intervals.

This paper presents an analogue of Dunnett's test procedure, a multivariate sign test. The data must consist of $(k+1)$-tuples, one observation on each of k treatments and one control, obtained under a variety of conditions, possibly quite different. The sign test for two treatments, for example a control and one treatment, is described by Dixon and Mood [1].

2. PROCEDURE

Let X_{oj} and X_{ij}, $i=1, \cdots, k$ and $j=1, \cdots, n$, be measured responses on the control and i-th treatment in the j-th block. The proposed multivariate sign test requires the number of + signs or − signs in each of the k sets of n signed differences between control and treatment.

The null and alternate hypotheses are stated in terms of the location of the medians of the multivariate distributions of the k-tuples of differences $(d_{1j}, \cdots, d_{kj})$ where we define median $(d_{1j}, \cdots, d_{kj})$ to equal (median $d_{1j}, \cdots,$ median d_{kj}). A common hypothesis is that the distributions of the $(d_{1j}, \cdots, d_{kj})$ have zero medians.

Let us suppose we wish to test each and every treatment against control for the purpose of locating treatments that give significantly greater responses than control. Then for a procedure using signed differences, the null and alternate hypotheses are:

H_o: Each k-tuple of differences $(d_{1j}, \cdots, d_{kj}) = (X_{1j} - X_{oj}, \cdots, X_{kj} - X_{oj})$, has a probability distribution with median zero.

H_1: The k-tuples of differences have probability distributions with common median in which one component is greater than zero.

The procedure for testing follows:

(1) Compute the signed differences $X_{ij} - X_{oj}$, $i = 1, \cdots, k$ and $j = 1, \cdots, n$.

(2) Observe the number of $-$ signs for each of the k sets of n signs and record as r_i, $i = 1, \cdots, k$.

(3) To judge significance, compare each r_i with the single tabulated critical value for the desired joint probability level. A significance statement is made for each of the k comparisons. The appropriate critical region is one tailed.

Application of the procedure for the purpose of locating treatments that give significantly smaller responses than control is obvious.

In case H_1 calls for a component of the common median to be simply different from zero (response significantly different from control), step 2 becomes:

(2′) Observe the number of times the less frequent sign occurs for each of the k sets of n signs and record as r_i, $i = 1, \cdots, k$.

Step 3 remains the same. However, the appropriate critical region is two-tailed.

Note that small values of r_i are declared significant. In other words, a value as small as or smaller than the tabulated value is declared significant.

The joint error rate for the test procedure is an experiment-wise or family-wise error rate. It is defined as the proportion of experiments in which at least one wrong inference is made when H_o is true.

An experiment-wise error rate makes us highly cautious in experiments with large numbers of treatments. This suggests that the significance level might be chosen according to the number of treatments, being larger as this number increases. Tables 772a and 772b give a limited number of complete distributions. Tables 769 and 770 are for significance levels of .05 and .01, those customarily used with per comparison error rates.

Ties have not been considered here although they will occur in practice. Only ties between treatment and check are of concern in this test. If an even number is present in any comparison, assign one-half this number to the appropriate r_i. If an odd number of ties is present, assign one at random, or in a manner dependent upon the penalty of a wrong decision, and the remainder equally as for an even number.

The usual modifications of the sign test may also be carried out for this multivariate sign test. In particular, we may test the hypotheses that the distributions of the k-tuples $(X_{oj} - a_1 X_{1j}, \cdots, X_{oj} - a_k X_{kj})$ or of the k-tuples $(X_{oj} - (A_1 + X_{1j}), \cdots, X_{oj} - (A_k + X_{kj}))$ have zero medians, testing for percentage or additive increases respectively.

3. EXAMPLE

The accompanying data are a small part of the results of the Cooperative Uniform Soybean Tests, 1956, for the North Central States. They consist of

TABLE 769. VALUES OF MINIMUM r FOR COMPARISON OF k TREATMENTS AGAINST ONE CONTROL IN n SETS OF OBSERVATIONS: ONE-TAILED CRITICAL REGION

Joint P	n	k = number of treatment means (excluding control)							
		2	3	4	5	6	7	8	9
.95	5	—	—	—	—	—	—	—	—
.99		—	—	—	—	—	—	—	—
.95	6	0	—	—	—	—	—	—	—
.99		—	—	—	—	—	—	—	—
.95	7	0	0	0	0	—	—	—	—
.99		—	—	—	—	—	—	—	—
.95	8	0	0	0	0	0	0	0	0
.99		—	—	—	—	—	—	—	—
.95	9	1	0	0	0	0	0	0	0
.99		0	—	—	—	—	—	—	—
.95	10	1	1	1	0	0	0	0	0
.99		0	0	0	0	—	—	—	—
.95	11	1	1	1	1	1	1	1	0
.99		0	0	0	0	0	0	0	0
.95	12	2	1	1	1	1	1	1	1
.99		1	0	0	0	0	0	0	0
.95	13	2	2	2	1	1	1	1	1
.99		1	1	1	0	0	0	0	0
.95	14	2	2	2	2	2	2	2	1
.99		1	1	1	1	1	1	0	0
.95	15	3	3	2	2	2	2	2	2
.99		2	1	1	1	1	1	1	1
.95	16	3	3	3	3	2	2	2	2
.99		2	2	1	1	1	1	1	1
.95	17	4	3	3	3	3	3	3	3
.99		2	2	2	2	2	1	1	1
.95	18	4	4	3	3	3	3	3	3
.99		3	2	2	2	2	2	2	2
.95	19	4	4	4	4	4	3	3	3
.99		3	3	2	2	2	2	2	2
.95	20	5	4	4	4	4	4	4	4
.99		3	3	3	3	3	2	2	2

TABLE 770. VALUES OF MINIMUM r FOR COMPARISON OF k TREATMENTS AGAINST ONE CONTROL IN n SETS OF OBSERVATIONS: TWO-TAILED CRITICAL REGION

Joint P	n	k = number of treatment means (excluding control)							
		2	3	4	5	6	7	8	9
.95	6	—	—	—	—	—	—	—	—
.99		—	—	—	—	—	—	—	—
.95	7	0	—	—	—	—	—	—	—
.99		—	—	—	—	—	—	—	—
.95	8	0	0	0	—	—	—	—	—
.99		—	—	—	—	—	—	—	—
.95	9	0	0	0	0	0	—	—	—
.99		—	—	—	—	—	—	—	—
.95	10	1	0	0	0	0	0	0	0
.99		0	—	—	—	—	—	—	—
.95	11	1	1	0	0	0	0	0	0
.99		0	0	0	—	—	—	—	—
.95	12	1	1	1	1	0	0	0	0
.99		0	0	0	0	0	—	—	—
.95	13	2	1	1	1	1	1	1	1
.99		0	0	0	0	0	0	0	0
.95	14	2	2	1	1	1	1	1	1
.99		1	1	0	0	0	0	0	0
.95	15	2	2	2	2	1	1	1	1
.99		1	1	1	1	0	0	0	0
.95	16	3	2	2	2	2	2	2	2
.99		1	1	1	1	1	1	1	1
.95	17	3	3	2	2	2	2	2	2
.99		2	2	1	1	1	1	1	1
.95	18	3	3	3	3	2	2	2	2
.99		2	2	2	1	1	1	1	1
.95	19	4	3	3	3	3	3	3	3
.99		2	2	2	2	2	2	1	1
.95	20	4	4	3	3	3	3	3	3
.99		3	2	2	2	2	2	2	2

yields in bushels per acre from two tests in Ontario, three in Ohio, one in Michigan, two in Wisconsin, two in Minnesota, two in North Dakota, and one in South Dakota. C is considered to be the standard or control variety. Clearly, the data were obtained under widely differing conditions.

Strain	Mean Yield in Bushels per Acre													
	Location													
	A	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*	*K*	*L*	*M*	Number of Minuses
X	29.2 (+)	21.4 (−)	36.3 (+)	40.7 (+)	39.2 (+)	45.6 (−)	20.5 (−)	26.2 (−)	34.4 (+)	46.1 (+)	6.0 (−)	19.8 (−)	24.0 (+)	6
Y	33.8 (+)	29.3 (+)	23.9 (+)	33.3 (−)	37.4 (+)	46.4 (−)	28.4 (+)	30.3 (+)	32.5 (−)	47.1 (+)	10.0 (+)	25.7 (−)	20.2 (−)	5
Z	31.3 (+)	29.5 (+)	24.4 (+)	30.8 (−)	37.4 (+)	43.5 (−)	28.4 (+)	29.8 (+)	33.5 (+)	44.5 (+)	9.0 (+)	29.1 (+)	24.5 (+)	2
C	23.8	25.4	17.2	33.5	34.9	49.4	24.2	28.4	32.8	44.4	8.5	27.3	20.8	

(These data are used with approval of the Field Crops Research Branch, ARS, USDA, and cooperating agencies.)

Reference to Table 769 shows that Z is significantly better than C, the tabulated value of min r for $\alpha = .05$ being 2.

4. DISTRIBUTION OF $(r_1, \cdots, r_k)$ AND MIN r_i

Consider the $(k+1)$-tuple that constitutes a single observation. Record the differences $X_{ij} - X_{oj}$, $i = 1, \cdots, k$, as 1 if negative and 0 if positive. This gives a vector of k components, each of which may be a 1 or a 0. The sum of the vectors gives the value of the test criterion $(r_1, \cdots, r_k)$, any component being the total number of minuses observed for the particular comparison.

For any trial, there are $(k+1)!$ equally likely arrangements, when the null hypothesis is true, of the $(k+1)$ observations. These give rise to 2^k possible vectors. Thus if $k=4$, there are $5! = 120$ possible arrangements but only $2^4 = 16$ possible vectors. The vector (1, 1, 1, 1) appears in $4! = 24$ possible arrangements (X_{oj} is the largest observation); the vector (1, 1, 0, 0) appears in $2!2! = 4$ possible arrangements; and so on. The probability with which any vector appears is the ratio of the product of the number of arrangements of the observations on each side of the control to the total number of arrangements.

Denote the set of possible vectors in a single trial by $v_1, \cdots, v_s$ where $s = 2^k$ and their probabilities by $p_1, \cdots, p_s$. Then the probability of obtaining the outcome $(r_1, \cdots, r_k)$ as the sum of the vectors in n trials is the sum of the coefficients of one or more terms in the expansion of expression (1).

TABLE 772a. EXACT PROBABILITY DISTRIBUTIONS FOR MINIMUM r_i
$k=2$, $n=4(1)10$

minimum r is equal to	Probability of event in column 1 for $n=$						
	4	5	6	7	8	9	10
0	.113	.058	.030	.015	.008	.004	.0019
1	.364	.251	.161	.098	.058	.033	.0188
2	.375	.381	.327	.253	.181	.122	.0793
3	.136	.244	.313	.326	.296	.245	.1878
4	.012	.062	.141	.221	.273	.289	.2716
5		.004	.027	.076	.141	.204	.2461
6			.001	.011	.038	.083	.1386
7				.000	.005	.018	.0466
8					.000	.002	.0086
9						.000	.0007
10							.0000

$$(p_1x_1 + \cdots + p_sx_s)^n \tag{1}$$

To find a particular term, first solve equation (2) for n_i's, subject to the restriction $\sum n_i = n$.

$$\sum n_iv_i = (r_1, \cdots, r_k) \tag{2}$$

The resulting solutions determine the appropriate terms in the expansion of expression (1) in that each solution is also a set of exponents of the x_i's and, hence, gives a term.

Consider the problem of computing the probability associated with a particular value of $(r_1, \cdots, r_k)$. For this, we first solve equation (2). The procedure for solution will now be illustrated for $k=3$, $n=6$ and the right side equal to (5, 4, 1); generalization of the procedure is obvious. Note that $k=3$, hence $s=2^3=8$. Write equation (2) as equation (3) including the restriction $\sum n_i = n$.

TABLE 772b. EXACT PROBABILITY DISTRIBUTIONS FOR MINIMUM r_i
$k=3$, $n=4(1)7$

Minimum r is equal to	Probability of event in column 1 for $n=$			
	4	5	6	7
0	.154	.082	.0430	.0221
1	.424	.326	.2114	.1339
2	.336	.389	.3682	.3051
3	.082	.184	.2775	.3263
4	.004	.018	.0894	.1692
5		.001	.0104	.0398
6			.0002	.0035
7				.0001

$$(n_1, \cdots, n_8)\begin{pmatrix}1111\\1101\\1011\\1001\\0111\\0101\\0011\\0001\end{pmatrix} = (5, 4, 1, 6) \tag{3}$$

Single trial vectors serve as the first k elements of the row vectors in the coefficient matrix and are ordered lexicographically. It is the latter fact that makes the procedure for obtaining solutions easy. Begin with the first column in the coefficient matrix and the restriction. Since $r_1=5$ and $n=6$, we need 5 ones and 1 zero. See step 1 in the accompanying scheme. This part-solution is carried into step 2 where the second column of the coefficient matrix is introduced. The required 4 ones may be obtained in two ways. These two part-solutions are now carried into step 3 where each gives three solutions. Notice that when any part solution includes a zero, there is no need to carry the corresponding vector into the next step. If, at each step, the solutions are obtained in an orderly fashion, there is little chance of missing or repeating one.

To compute probabilities, we now return to the expansion of expression (1). Probabilities p_i are computed as described in paragraph 2 of this section. For $k=3$, there are $(3+1)!=24$ equally likely arrangements of the observations. Where X_o is the least or greatest observation, there are $3!=6$ arrangements giving the same vector; where X_o is not least or greatest, there are $2!=2$ arrangements that give the same vector. Thus for equation (3) the first and last solutions have probabilities

$$\frac{6!}{3!}\,\frac{6^2 2^4}{(24)^6} \quad \text{and} \quad \frac{6!}{3!2!}\,\frac{2^6}{(24)^6}\cdot$$

The probability that $(r_1, r_2, r_3)=(5, 4, 1)$ is the sum of the six probabilities so computed. This probability applies to each of 12 vectors, those with numbers which are the 6 permutations of 5, 4, 1 and those with numbers which are the 6 permutations of 1, 2, 5, the numbers in the complement of (5, 4, 1). Because of symmetry, $(r_1, \cdots, r_k)$ and its complement $(n-r_1, \cdots, n-r_k)$ have the same probability of occurrence. Note that the probability associated with the vector (5, 3, 1) applies to only 6 vectors because the numbers in the complement of (5, 3, 1) are simply a permutation of the same numbers.

When probabilities have been computed for the minimum number of terms necessary to construct the complete probability distribution, the sum of the products of the probabilities and the number of terms having the specified probability serves as a check on the procedure. (Unfortunately, the number of terms to be computed increases rapidly as either k or n increases.) From this distribution, the distribution of the minimum r_i is obtained. This is the required distribution.

SCHEME FOR SOLVING EQUATION 3

Step 1

	Part solution
1	5
0	1
	6

Step 2

			Part solution	
	1	1	4	3
	1	0	1	2
from step 1			5	5
	0	1	0	1
	0	0	1	0
from step 1			1	1
			6	6

Step 3

			Solutions					Solutions		
	11	1	1	0	0	11	1	1	0	0
	11	0	3	4	4	11	0	2	3	3
from step 2			4	4	4			3	3	3
	10	1	0	1	0	10	1	0	1	0
	10	0	1	0	1	10	0	2	1	2
from step 2			1	1	1			2	2	2
	00	1	0	0	1	01	1	0	0	1
	00	0	1	1	0	01	0	1	1	0
from step 2			1	1	1			1	1	1
			6	6	6			6	6	6

5. AN APPROXIMATION

An obvious conjecture is that $(r_1, \cdots, r_k)$ is from a multivariate normal distribution and that $(\min r-\mu_r)/\sigma_r$ is distributed approximately as Dunnett's [2] t for infinite degrees of freedom. For this approximation, $\mu_r=n/2$, $\sigma_r^2=n/4$. However, Dunnett's t is computed on the basis that $\rho=0.5$, whereas for the distribution of $(r_1, \cdots, r_k)$, the correlation between r_i and r_j is $\rho=\frac{1}{3}$. Roessler [3] has computed tables with $\rho=0$, which are comparable to Dunnett's tables for two-sided comparisons and joint confidence coefficients of $P=.95$ and .99. A comparison shows that corresponding tabulated values differ only in the second decimal place for $P=.95$ and never by more than .1 for $P=.99$. (Dunnett's table gives two decimal places, Roessler's gives one place.) Since the appropriate ρ lies between those used by Roessler and Dunnett, since the Roessler and Dunnett tables differ so little, and since Dunnett gives two decimal places, it was decided to use the latter in computing tables.

Tables 769 and 770 were computed by taking the integral part of the number computed by means of equation 4 with t from Dunnett's tables. Where the computation gave negative values, it was assumed that no value of r_i should be declared significant. The equation was suggested by approximations given by Dixon and Mood [1], the final form being chosen as a result of comparing computed values with the exact values obtainable from Tables 772a and 772b.

$$r = \frac{n-1}{2} - t\frac{\sqrt{n}}{2} \qquad \text{(Dunnett's } t\text{)} \tag{4}$$

Of the 32 comparable values, 4 differed, the approximation giving no value as significant whereas $r=0$, was significant for the first time with increasing n. These discrepancies occurred for $k=2$, $n=8$, $P=.01$ (one tail), for $k=3$, $n=6$,

$P = .05$ (one tail), for $k = 2$, $n = 9$, $P = .01$ (two tails), and for $k = 3$, $n = 7$, $P = .05$ (two tails), where equation 4 gave $-.008$, $-.021$, $-.185$ and $-.136$ respectively. The corresponding true probabilities for $r = 0$ were .008, .0430, $\leq .008$ and $\leq .0442$.

REFERENCES

[1] Dixon, W. J. and Mood, A. M., "The statistical sign test," *Journal of the American Statistical Association*, 41 (1946), 557–66.

[2] Dunnett, C. W., "A multiple comparison procedure for comparing several treatments with a control," *Journal of the American Statistical Association*, 50 (1955), 1096–121.

[3] Roessler, E. B., "Testing the significance of observations compared with a control," *Proceedings of the American Society for Horticultural Science*, 47 (1946), 249–51.