

MULTIPLE RANGE TESTS FOR CORRELATED AND HETEROSCEDASTIC MEANS

by

David B. Duncan

Universities of Florida and North Carolina

This research was supported jointly by the Florida Agricultural Experiment Station, by the U. S. Public Health Service and by the United States Air Force through the Office of Scientific Research of the Air Research and Development Command.

Institute of Statistics  
Mimeograph Series No. 161  
December, 1956

MULTIPLE RANGE TESTS FOR CORRELATED AND HETEROSCEDASTIC MEANS.\*

by

David B. Duncan

Universities of Florida and North Carolina

1. INTRODUCTION

Multiple range tests have been developed by several writers, for example, D. Newman [8], M. Keuls [5], J. W. Tukey [10] and D. B. Duncan [3], for testing differences between several treatment means in cases in which all such differences are of equal a priori interest. These tests, which are also described in recent textbooks, for example, W. T. Federer [4, chapter 2], have been worked out for data in which the treatment means are homoscedastic (have equal variances) and are uncorrelated. Recently, C. Y. Kramer [6] has presented a simple method for extending these procedures to give useful tests for differences between means with unequal replications, the method being applicable to any set of heteroscedastic uncorrelated means. In a subsequent paper [8], the same author has given further extensions to tests of means which are also correlated, such as the adjusted means from analyses of covariance or from incomplete block designs. Similar work has also been done by E. Bleicher [1] and P. G. Sanders [9] in extending a multiple F test to making tests in lattice and rectangular lattice designs.

One purpose of this paper is to present a more complete method for these extensions which necessarily sacrifices a little in simplicity but is more powerful, especially in cases in which the differences between the means have appreciably different variances. Another purpose is to indicate briefly the closeness of the properties of these complete tests of heteroscedastic and correlated means to those of the corresponding tests of homoscedastic and uncorrelated means. Incidental to these main purposes, a short-cut skipping principle, useful in applying multiple range tests to a large number of treatment means (or totals), is also presented.

---

\*Research jointly supported by the Florida Agricultural Experiment Station, by the U. S. Public Health Service and by the U. S. Air Force through the Office of Scientific Research of the Air Research and Development Command.

## 2. BASIC RULE FOR COMPLETE TESTS

Let  $m_1, m_2, \dots, m_n$  represent  $n$  normally distributed means such that the variance of the difference between each pair can be written  $V(m_i - m_j) = k_{ij}\sigma^2$  where  $k_{ij}$  is known and  $\sigma^2$  is an expected error mean square. Let  $s^2$  with  $n_2$  degrees of freedom be the usual type of analysis of variance error mean square estimate for  $\sigma^2$ . In other words,  $n_2 s^2 / \sigma^2$  is distributed as  $\chi^2_{n_2}$  and is independent of the means  $m_1, m_2, \dots, m_n$ .

Call  $a_{ij} = \sqrt{2/k_{ij}}$  the adjustment factor for and  $(m_i - m_j)' = a_{ij}(m_i - m_j)$  the adjusted difference between the means  $m_i$  and  $m_j$ , and call  $R'_p = s \cdot z_p$  the critical value for  $p$  means, where  $z_p$  is the Studentized significant range for  $p$  means for  $n_2$  degrees of freedom and for an  $\alpha$ -level test.

The proposed complete basic rule for an  $\alpha$ -level multiple range test may then be expressed as follows: Any subset of  $p$  means is homogeneous if the largest adjusted difference in the subset fails to exceed the critical value  $R'_p$ . Any two means not both contained in the same homogeneous subset are significantly different. Any two means both contained in the same homogeneous subset are not significantly different.

This is the same as the basic rule implicitly adopted by Kramer [6] except that in the latter a subset of  $p$  means is declared homogeneous if its adjusted range does not exceed  $R'_p$ . If an adjusted difference within a subset exceeds the adjusted range, as it may do through having a smaller variance and hence a larger adjustment factor, it will be significant by the complete rule and this may also result in the detection of further significant differences.

## 3. NUMERICAL EXAMPLE I: TEST OF UNEQUALLY-REPLICATED MEANS.

Table 1 illustrates a convenient method for applying the complete rule. The example consists of the application of a 5 per cent level new multiple range test [3] to a set of seven unequally-replicated treatment means from a completely randomized design. A similar extension of any multiple range test, e.g., [5], [8], and [10], could be made by the same method, the only difference being in the

source used for the Studentized significant ranges  $z_p$ , in section (b). Table 2 gives details of the computation of the adjusted differences used in Table 1.

The initial preparation of the data is the same as for Kramer's method [6]. Table 1, section (a) shows the analysis of variance concluding with the calculation of the error standard deviation  $s = 73.45$ . Section (b) shows the computation of the critical values  $R'_p = s \cdot z_p$ , the Studentized significant ranges  $z_p$  having been taken from [3, Table II] for a 5 per cent level test entering at the row for  $n_2 = 16$  degrees of freedom. Section (c) shows the treatment means ranked in ascending order together with their respective replication numbers in parentheses. In any test of uncorrelated means it is helpful to list under the ranked means measures, such as replication numbers in this case, which provide a quick method of visually assessing the relative magnitudes of the variances of the means and hence of the variances of the differences between them.

The main part of the test is in the sequences of steps in section (d). Each step consists of an application of the basic rule to a particular subset. Sequence 1 consists of steps involving all subsets in which the top mean G is the largest mean, sequence 2 involves all subsets in which the second mean E is the largest mean, and in general, sequence  $i$  involves all subsets in which the  $i$ -th mean is the largest mean.

The order of steps in each sequence is the same at the beginning as that of previous procedures [3] and [6]. The steps in sequence 1, for example, consist of testing the adjusted ranges first of the whole set DFABCEG, then of the subset FABCEG, then of ABCEG and so on. At each step the lowest mean is dropped to give the subset for the next test.

TABLE 1. 5 Per Cent Level New Multiple Range Test [3] of Seven Unequally-Replicated Means

a) Analysis of Variance

<u>Source</u>	<u>d.f.</u>	<u>m.s.</u>	<u>s = <math>\sqrt{m.s.}</math></u>
Between treatments	6		
Error	16	5,394.6	73.45

b) Critical Values:  $R'_p = s \cdot z_p$

	(2)	(3)	(4)	(5)	(6)	(7)
$z_p$ :	3.00	3.15	3.23	3.30	3.34	3.37
$R'_p$ :	220.4	231.4	237.2	242.4	245.3	247.5

c) Ranked Treatment Means and Replication Numbers

D	F	A	B	C	E	G
680	734	743	851	873	902	945
(3)	(2)	(5)	(5)	(3)	(2)	(3)

d) Test Sequences

<u>Seq.</u>	<u>Steps</u>	<u>Result</u>
1.	$(G-D)' > R'_7, (G-F)' > R'_6, (G-A)' > R'_5, (G-B)' \not> R'_4$	(BCEG)
2.	$(E-D)' > R'_6, (E-F)' \not> R'_5, FABCE: (E-A)' > R'_5, FBCE: (E-F)' > R'_4, BCE.$	-
3.	$(C-D)' > R'_5, (C-F)' \not> R'_4, FABC: (C-A)' > R'_4, FBC: (C-F)' \not> R'_3, (C-B)' \not> R'_3, (B-F)' \not> R'_3.$	(FBC)
4.	$(B-D)' > R'_4, (B-F)' \not> R'_3, FAB: (B-A)' > R'_3, FB.$	-
5.	$(A-D)' \not> R_3.$	(DFA)

e) Final Result

(DFA) (FBC) (BCEG)

Any two means not appearing together within the same parentheses are significantly different. Any two means appearing together within the same parentheses are not significantly different.

TABLE 2. Arithmetical Details for Calculating Adjusted Differences in Table 1, section (d).

$$a_{1j} = \sqrt{2/k_{1j}} = \sqrt{2/(\frac{1}{r_1} + \frac{1}{r_j})} = \sqrt{2r_1r_j/(r_1+r_j)}, \text{ where } r_1, r_2, \dots, r_7 \text{ are}$$

the replication numbers for the respective means, thus  $(G-D)' = (G-D)a_{GD} =$

$$265 \sqrt{2(3)(3)/6} = 265(1.732) = 459.0. \text{ Similarly}$$

$(G-F)'$	$= 211(1.549)$	$= 326.8,$	$(G-A)'$	$= 202(1.936)$	$= 391.1,$
$(G-B)'$	$= 94(1.936)$	$= 182.0,$	$(E-D)'$	$= 222(1.549)$	$= 343.9,$
$(E-F)'$	$= 168(1.414)$	$= 237.6,$	$(E-A)'$	$= 159(1.690)$	$= 268.7,$
$(C-D)'$	$= 193(1.732)$	$= 334.3,$	$(C-F)'$	$= 139(1.549)$	$= 215.3,$
$(C-A)'$	$= 130(1.936)$	$= 251.7,$	$(C-B)'$	$= 22(1.936)$	$= 42.6,$
$(B-F)'$	$= 117(1.690)$	$= 197.7,$	$(B-D)'$	$= 171(1.936)$	$= 331.1,$
$(B-A)'$	$= 108(2.236)$	$= 241.5,$	$(A-D)'$	$= 63(1.936)$	$= 122.0.$

The changes in the complete test come in each sequence when the adjusted range of a subset of  $p$  means fails to exceed  $R'_p$ . If the adjustment factor for the range is smaller than that of any other difference in the subset, or, in a test like this of unequally-replicated means, if either of the extreme means has fewer replications than any of the other means, any of the other differences with a larger adjustment factor, should also be tested. In such cases it is helpful to write down the subset concerned as a reminder that it will still be the subset under test until an adjusted difference within it is found to exceed  $R'_p$ . For example, when the adjusted range  $(C-F)'$  of FABC fails to exceed  $R'_4$  in step 2 of sequence 3, the subset FABC is written down before testing the adjusted difference  $(C-A)'$  in the next step. This serves as a reminder that  $(C-A)'$  must be compared with  $R'_4$  and not  $R'_3$  as would otherwise have happened. Similarly in the next three steps, the preliminary recording of FBC serves as a reminder that  $(C-F)'$ ,  $(C-B)'$  and  $(B-F)'$  each have to be compared with  $R'_3$ .

When an adjusted difference between the top mean in a subset of  $p$  means and an internal mean is found to exceed  $R'_p$ , the internal mean is dropped to give the next subset to be tested. For example, in step 3 of sequence 3 FABC is reduced to FBC in this way by dropping A when  $(C-A)'$  exceeds  $R'_4$ .

When an adjusted difference, not involving the top mean, is found to exceed  $R'_p$ , two subsets may qualify for further testing in the same sequence. For example, if four means were ranked and had replication numbers as follows

P	S	R	Q
(1)	(50)	(100)	(1)

the testing steps could be

$$(Q-P)' \not> R'_4, PSRQ: (Q-S)' \not> R'_4, (Q-R)' \not> R'_4, (R-P)' \not> R'_4, (R-S)' \not> R'_4,$$

and the subsets PRQ and PSQ would both qualify for testing in further steps in the same sequence.

The testing for a subset terminates either when it is shown to be homogeneous, which fact is recorded by noting the subset in parentheses in the result column at the end of the sequence involved, or when the subset is found to be completely included within another subset already shown to be homogeneous. For example, BCEG is recorded (BCEG) at the end of the first sequence to denote its homogeneity. This follows from the fact that the adjusted range  $(B-G)'$  of BCEG does not exceed  $R'_4$  and neither of the replications 5 and 3 of B and G is less than the replications 3 or 2 of C or E. The result (DFA) of sequence 5 is of a similar form. In other cases, e.g., the result (FBC) in sequence 3, it is sometimes necessary to test each adjusted difference in the subset before it can be declared homogeneous.

Sequences 2 and 4 provide no additional homogeneous subsets because they terminate at subsets BCE and FB, which are included in (BCEG) and (FBC) respectively already shown to be homogeneous.

It should be noted in conclusion that it is possible for more than one homogeneous subset to be found in a single sequence. For example, in the case of the

means PSRQ discussed above the sequence involved could terminate with the results (PRQ) and (PSQ) or even (PQ) (SQ) and (RQ) depending on the other data involved.

Section (e) of Table 1 shows a useful way of presenting the results of the test. The device of presenting the whole set with homogeneous subsets underscored as is done in [3] and [6] cannot be used here because of the differences in the order of the means in the various homogeneous subsets. For example, A is to the right of F in (DFA) but not in (FBC). The new method of putting homogeneous groups in parentheses can also be used in tests of equally replicated means and may be preferred for printing purposes.

#### 4. NUMERICAL EXAMPLE II: TEST OF TREATMENT TOTALS IN A SIMPLE LATTICE DESIGN

(Including the Use of Skipping Short Cuts).

Table 3 illustrates the application of a similar 5 per cent level test to the adjusted totals in a 5 x 5 simple lattice design. The data are those given by Cochran and Cox [2, section 10.29] for a design with two repetitions. Table 4 gives additional details of the computation of the adjusted differences used in Table 3.

Section (a), Table 3, shows the value of  $s$  obtained from the error mean square  $s^2$  (denoted  $E_e$  in [2]) for the experiment and its degrees of freedom  $n_2$ . Section (b) shows the adjustment factors for differences between treatment totals (totals being more convenient than means to use in a case like this).

Cochran and Cox give  $\frac{2E_e}{r} [1+(n-1)\mu]$ , (their  $n$  being the number (2) of repetitions involved) for the estimated variance of a difference between two means for treatments in the same block. Thus, using  $k_{++}\sigma^2$  to denote the variance of a difference between totals for treatments in the same block we have  $k_{++} = 2r [1+(n-1)\mu]$ . Then using  $a_{++}$  for the corresponding adjustment factor, we have  $a_{++} = \sqrt{2/k_{++}} = (r [1+(n-1)\mu])^{-(1/2)}$ . Similarly, if  $a_{+-}$  is used to denote the adjustment factor for differences between totals of treatments not in the same block, we have  $a_{+-} = \sqrt{2/k_{+-}} = (r [1 + n\mu])^{-(1/2)}$ . In this example,  $r = 4$ ,  $n = 2$ ,  $\mu = 0.1270$  and the adjustment factors work out to be as shown in section (b).

Section (c) shows the ranked treatment totals and critical values required for the test. The arrangement is different from the corresponding sections of the previous example solely because of the largeness of the number (25) of treatments involved. The new arrangement is convenient for applying a skipping method which short cuts many of the steps at the beginning of each sequence. In all other respects the procedure is unchanged. In the first column the treatments 1, 2, ..., 25 as they have been denoted in [2] are redenoted A, B, ..., Y for convenience in the recording of treatment subsets. The number (i.j) appearing after each treatment letter denotes the blocks in which the treatment falls. Thus (3.1) after treatment K shows that it belongs to block 3 and to block 1 in the first and second types of replicates, respectively. These numbers are useful in indicating which treatments do and which do not appear together in the same block and thus which adjustment factor applies to each difference.

The second column of section (c) shows the adjusted treatment totals from [2] followed by doubly adjusted treatment totals in parentheses which, for brevity, we will call treatment totals and adjusted treatment totals, respectively. Each adjusted treatment total in parentheses is obtained by multiplying the corresponding treatment total by the smallest adjustment factor. In this example there are only two adjustment factors, the smaller being .447, hence the adjusted total for K, for instance, is  $88.4(.447) = 39.5$  as shown. The column of adjusted totals is a new feature needed in the skipping short cut steps.

The last two columns of section (c) show the Studentized ranges  $z_p$  and the critical values  $R'_p = 3.69 z_p$ . The middle column for p helps in identifying the  $z_p$  and  $R'_p$  values. The  $z_p$  values in this example are obtained from [3, Table II] for a 5 per cent level test the same as in the previous example except that for larger values of p some simple linear interpolation is needed. When a large number of treatments is involved as in this example, not all of the critical values  $R'_p$  will be required. Each one should thus be obtained only as needed in the sequence steps. In Table 3, for instance, only 12 of the possible 24  $R'_p$  values are ultimately found to be needed.

Table 3. 5 Per Cent Level Multiple Range Test of Adjusted Treatment Totals from a 5 x 5 Simple Lattice Design.

a) From Analysis of Variance

$$n_2 = 56, \quad s^2 = 13.60, \quad s = 3.69$$

b) Adjustment Factors for Differences between Treatment Totals

Two treatments in same block:  $a_{++} = .471$

Two treatments not in same block:  $a_{+-} = .447$

c) Ranked Treatment Totals and Critical Values

<u>Treatment</u>	<u>Total</u>	<u>p</u>	<u><math>\frac{z}{p}</math></u>	<u><math>\frac{R'_p}{p}</math></u>
11 K (3.1)	88.4 (39.5)	25	3.47	12.80
2 B (1.2)	77.3 (34.6)	24	3.47	12.80
15 O (3.5)	74.7 (33.4)	23	3.47	12.80
14 N (3.4)	71.6 (32.0)	22	3.47	12.80
24 X (5.4)	70.6 (31.6)	21	3.47	12.80
22 V (5.2)	68.1 (30.4)	20	3.47	12.80
1 A (1.1)	66.6 (29.8)	19	3.46	12.77
21 U (5.1)	61.4 (27.4)	18	3.45	12.73
4 D (1.4)	58.8 (26.3)	17	3.44	12.69
16 P (4.1)	58.3 (26.1)	16	3.43	12.66
23 W (5.3)	55.7 (24.9)	15		
25 Y (5.5)	52.7 (23.6)	14		
13 M (3.3)	52.7 (23.6)	13		
18 R (4.3)	52.6 (23.5)	12		
20 T (4.5)	51.6 (23.1)	11		
12 L (3.2)	51.1 (22.8)	10		
5 E (1.5)	50.9 (22.8)	9		
7 G (2.2)	47.6 (21.3)	8	3.28	12.10
6 F (2.1)	46.9 (21.0)	7	3.25	11.99
10 J (2.5)	46.2 (20.7)	6		
17 Q (4.2)	46.0 (20.6)	5		
8 H (2.3)	45.2 (20.2)	4		
3 C (1.3)	44.9 (20.1)	3		
9 I (2.4)	38.1 (17.0)	2		
19 S (4.4)	21.5 ( 9.6)	-		

d) Test Sequences

<u>Seq.</u>	<u>Steps</u>	<u>Result</u>
1.	$39.5 - R'_{25} = 26.70, 39.5 - R'_8 = 27.4.$ $a_{++}(K - U) > R'_8, a_{++}(K - A) \not> R'_7.$	(AVXNOBK)
2.	$34.6 - R'_{24} = 21.80, 34.6 - R'_{16} = 21.94.$ $a_{++}(B - E) \not> R'_{16}.$	(ELTRMYWPDUA VXNOB)
3.	$33.4 - R'_{23} = 20.60, 33.4 - R'_{19} = 20.63, 33.4 - R'_{18} = 20.67.$ $a_{++}(O - J) > R'_{18}. FGEL...0: -$	(FGELTRMYWPDUA VXNO)
4.	$32.0 - R'_{22} = 19.20, 32.0 - R'_{20} = 19.20.$ CHQJ...N: -	(CHQJFGELTRMYWPDUA VXN)
5.	$31.6 - R'_{21} = 18.80.$	----
6.	$30.4 - R'_{20} = 17.60.$	----
7.	$29.8 - R'_{19} = 17.03.$	----
8.	$27.4 - R'_{18} = 14.67, -$	(ICHQJFGELTRMYWPDU)
last	$9.6 + R'_{17} = 22.29, 9.6 + R'_8 = 21.70,$ SICHQJFG: $a_{++}(Q - S) \not> R'_8.$	(SICHQJFG).

e) Final Results

(SICHQJFG)(ELTRMYWPDUA VXNOB)  
(ICHQJFGELTRMYWPDU)(AVXNOBK)  
(CHQJFGELTRMYWPDUA VXN)  
(FGELTRMYWPDUA VXNO)

Any two treatments not appearing together within the same parentheses are significantly different. Any two treatments appearing together within the same parentheses are not significantly different.

Table 4. Arithmetical Details for Calculating Adjusted Differences in Table 3, section (

$$(K - U)' = a_{++}(K - U) = .471(88.4 - 61.4) = .471(27.0) = 12.72$$

Similarly

$$(K - A)' = .471(21.8) = 10.27, (B - E)' = .471(26.4) = 12.43,$$

$$(O - J)' = .471(28.5) = 13.42, (Q - S)' = .471(24.5) = 11.54.$$

Section (d) shows the main part of the test arranged in steps within sequences as in the previous example. The first two steps in sequence 1 are skipping steps and short cut the individual testing of 17 differences. In the first step, the largest critical value,  $R'_{25} = 12.80$  is subtracted from the largest adjusted total 39.5 (for K) giving 26.70. From this it is concluded that all treatments with adjusted totals below 26.70, namely D, P, ..., S, are significantly lower than K. They can thus be dropped from the set leaving the subset UA...K for testing in the second step.

The truth of this conclusion is readily seen as follows: — Consider any one of the differences concluded significant, say K - M for example. We have  $M(.447) < K(.447) - R'_{25}$  implying  $(K - M)(.447) > R'_{25}$ , thus  $(K - M)(.471) > R'_{13}$ , since  $R'_{25} > R'_{13}$ , and hence  $(K - M)' > R'_{13}$ . Similarly each of the adjusted differences concerned exceeds its corresponding critical value and all are thus significant.

In the second step testing the 8-treatment subset UA...K in a similar way, the largest critical value  $R'_8$  involved is subtracted from the top adjusted total giving 27.4. If there were any further adjusted totals below 27.4 the treatments concerned could be dropped off and another similar step would be applied. In these data no further treatments can be dropped and the skipping method terminates at this second step. The remainder of the sequence is finished by steps of the type already described in the first example, and for which the arithmetical details are given in Table 4. Thus in step 3,  $a_{++}(K - U) > R'_8$ , and in step 4,  $a_{++}(K - A) > R'_7$ . This terminates the sequence since K - A has the larger adjustment factor  $a_{++}$  and no other difference within the subset AV...K can exceed  $R'_7$ . This result may be usefully recorded as before by putting the subset in parentheses as shown in the result column of section (d) at the end of sequence 1.

Sequence 2 is very similar to sequence 1. The skipping procedure starts by adding  $R'_{24}$  to the second highest adjusted total 34.6 for B and terminates again in the second step. The remainder of the sequence terminates at the third step with EL...B being found homogeneous.

In sequence 3 an additional treatment is dropped in the second step and the skipping procedure extends to the third step. Continuing the remainder of the sequence,  $a_{++}(O - J)$  is found to exceed  $R'_{18}$  in the fourth step,  $a_{+-}(O - F)$  is known not to exceed  $R'_{17}$  from the preceding skipping steps so FGEL...0 is recorded at the fifth for further internal testing. The largest (O - ?) difference with an  $a_{++}$  adjustment factor is O - E. However  $a_{++}(O - E)$  cannot exceed  $R'_{17}$  since  $a_{++}(B - E) \not> R'_{16}$  in sequence 2 hence FG...0 is homogeneous and the sequence terminates. Sequence 4 has only two skipping steps and terminates in a similar way. Sequences 5, 6 and 7 each terminate rapidly at the first step when the subsets concerned are found to fall entirely within (CH...N) of sequence 4. In sequence 8, the difference between the adjusted total for I, that is, 17.0 and the critical level 14.67 is so great as to leave no doubt of the final result (IC...U) at the end of the first step.

As soon as all treatments but the lowest have been included in homogeneous subsets, as is the case in the example at the end of sequence 8, the test can be completed in one reverse type of sequence working from the bottom total. The reasons for this will be evident from the steps of the last sequence in section (d). The largest possible homogeneous subset in which the bottom treatment S could be included at this stage is SI...D and contains 17 treatments. The first step is to obtain  $S + R'_{17} = 9.6 + 12.69 = 22.29$ . From this it follows that all treatments with adjusted totals above 22.29, that is E, L, T, ... are significantly larger than S. This leaves the 8-treatment subset SI...G for testing in the next step. In the second step  $S + R'_8 = 21.70$ , no additional adjusted totals exceed this and the skipping procedure terminates. Since the range G - S of the subset has the adjustment factor  $a_{+-}$  we already know from the step 2 that  $a_{+-}(G - S) \not> R'_8$  hence SICH0JFG is recorded for internal testing. The largest (? - S) difference with the  $a_{++}$  adjustment factor is (Q - S) and this is therefore tested in step 3. Since  $a_{++}(Q - S) \not> R'_8$  the subset is homogeneous, and is recorded (SICH0JFG). This terminates the test.

Section (e) of Table 3 shows the complete summary of the test results. In this example the ordering of treatments does not vary from one homogeneous subset to another. In such a case the method of representing the results by underscoring a single set of treatment letters as is done in [3] may be used if preferred.

### 5. NOTES ON THE PROPERTIES OF THE PROPOSED TEST

Two-mean significance levels: A two-mean significance level in a test of  $n$  means may be defined [3] as the maximum probability of finding a significant difference between any two means  $m_i$  and  $m_j$  given that  $\mu_i = \mu_j$  where  $\mu_i = E(m_i)$  and  $\mu_j = E(m_j)$ . This may be written as the  $\max. P[D_{ij} | \mu_i = \mu_j]$  where  $D_{ij}$  denotes the decision that  $m_i$  and  $m_j$  are significantly different.

In any  $\alpha$ -level test of the proposed type we have  $\max. P[D_{ij} | \mu_i = \mu_j] = P[a_{ij} | m_i - m_j > s \cdot z_{2, n_2, \alpha} | \mu_i = \mu_j]$ . Since  $z_{2, n_2, \alpha} = \sqrt{2} t_{n_2, \alpha}$  where  $t_{n_2, \alpha}$  the  $\alpha$ -level (two-sided) significant value of  $t_{n_2}$ , a  $t$  statistic with  $n_2$  degrees of freedom, and since the variance of  $m_i - m_j$  is  $2\sigma^2/a_{ij}^2$ , this readily reduces to

$$P[|t_{n_2}| > t_{n_2, \alpha}] = \alpha.$$

Hence the two-mean significance levels in an  $\alpha$ -level test of the proposed type are exactly  $\alpha$  as desired.

Higher order significance levels and power: In considering these further aspects of the proposed test it is helpful to study the decision regions for a 5 per cent level test [3] of three unequally replicated means  $m_1$ ,  $m_2$  and  $m_3$  with  $r_1 = 2$ ,  $r_2 = 3$  and  $r_3 = 4$  replications and in which  $n_2 = \infty$  and  $s^2 = \sigma^2 = 1$ . If these regions are plotted in a plane with coordinates  $x_1 = (m_1 - m_2) \sqrt{2r_1 r_2 / (r_1 + r_2)}$  and

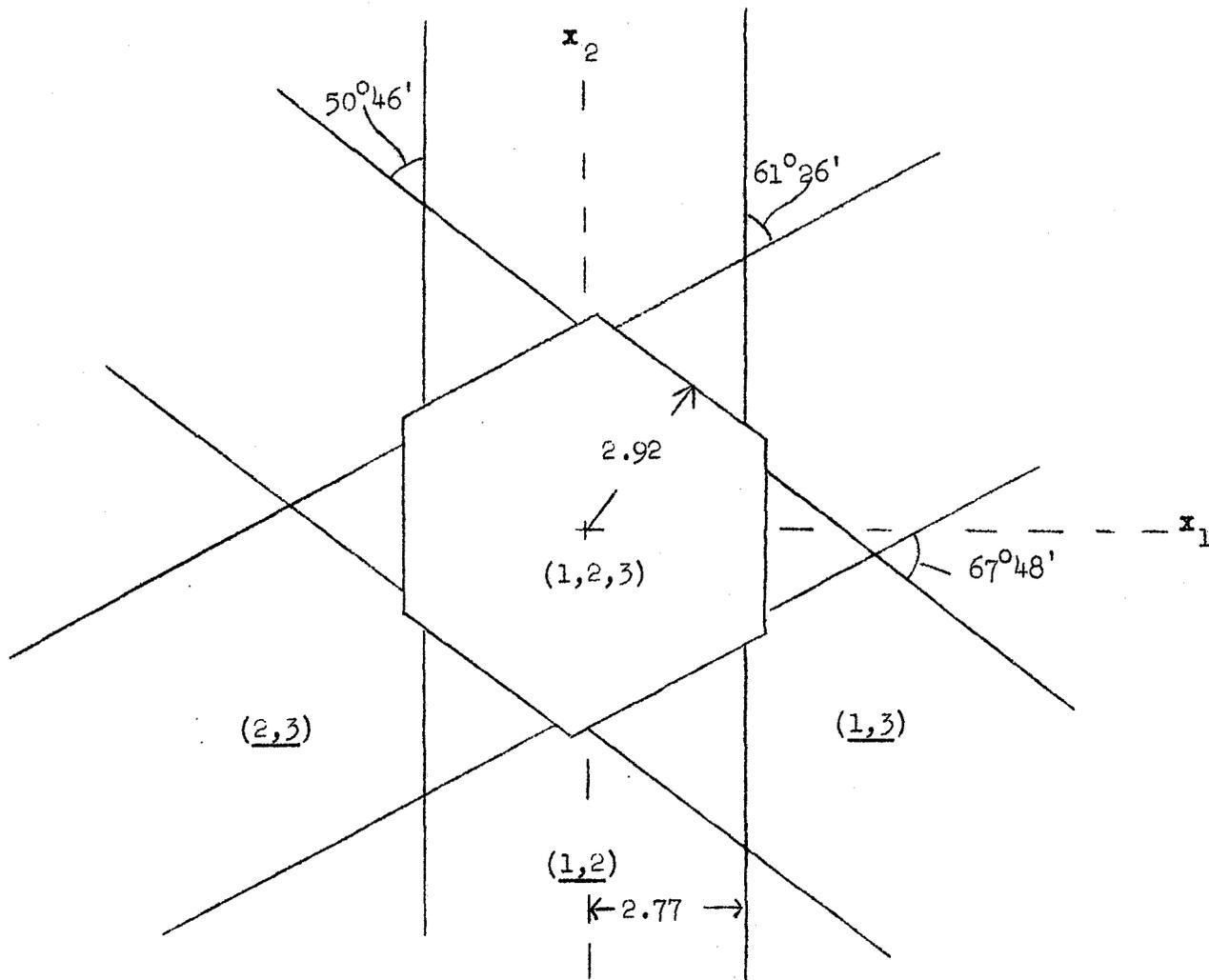


Figure 1

5% level test,  $r_1 = 2, r_2 = 3, r_3 = 4, n_2 = \infty, \sigma^2 = 1.$

$x_2 = (r_1 m_1 + r_2 m_2 - (r_1 + r_2) m_3) \sqrt{2r_3 / (r_1 + r_2)(r_1 + r_2 + r_3)}$  as is done in Figure 1, they are directly comparable to those of the corresponding 5 per cent level test of three equally replicated means ( $r_1 = r_2 = r_3$ ) shown in Figure 3 of [3]. The distribution function for the points  $(x_1, x_2)$  is the same in both cases, namely a bivariate normal with variances 2 and 2 and with covariance zero.

Because the strip regions  $(1, 2)$ ,  $(1, 3)$  and  $(2, 3)$  have the same minimum widths in each case ( $2z_{2,\infty} = 2 \times 2.77$ ) it follows, as has already been proved, that the two-mean significance levels are 5 per cent for the test in Figure 1 as well as for the test with equal replications.

The only differences between the regions in the two cases is that in Figure 1 the angles between the strip regions  $(1, 2)$ ,  $(1, 3)$ ,  $(2, 3)$  and  $(1, 2)$  are  $50^\circ 46'$ ,  $67^\circ 48'$  and  $61^\circ 26'$  instead of all being  $60^\circ$  as in the other figure. (The cosine of the angle between any two strips  $(h, i)$  and  $(h, j)$  is given by

$$\sqrt{r_i r_j / [(r_h + r_i)(r_h + r_j)]}.$$

The sides of the hexagonal regions  $(1, 2, 3)$  are parallel with the corresponding strips. Since these hexagons have the same inscribed circle of radius  $z_{3,\infty} = 2.92$  and differ only in having a little asymmetry in Figure 1 the three-mean protection level (the probability  $P[(x_1, x_2) \in (1, 2, 3) | E(x_1) = E(x_2) = 0]$ ) for the Figure 1 test is a close approximation to the desired level .9025 obtained in the other test. Furthermore, it seems safe to assume that any deviation due to the asymmetry would be positive.

In terms of three-mean significance levels, the level of the Figure 1 test may thus be said to be close to .0975 ( $= 1 - .9025$ ) as desired and that any deviation from this appears to be on the negative or conservative side.

The close similarity of the regions of Figure 1 with those of Figure 3 [3] also indicates that the power functions of the Figure 1 test closely approximate the

desirable ones of the other procedure.

If the means are correlated as well as being heteroscedastic the geometrical picture is virtually unchanged. If we let  $\begin{bmatrix} c_{ij} \end{bmatrix}_{3 \times 3}$  represent the dispersion matrix in the case of three means the cosine of the angle between any two strips  $(h,i)$  and  $(h,j)$  in a set of regions otherwise similar to those of Figure 1 would be

$$(c_{hh} - c_{hi} - c_{hj} + c_{ij}) / \sqrt{(c_{hh} - 2c_{hi} + c_{ii})(c_{hh} - 2c_{hj} + c_{jj})}$$

This may be expressed in terms of the  $k_{ij}$  factors ( $k_{ij}^2 = c_{ii} - 2c_{ij} + c_{jj}$ ) defined in section 2, as

$$(k_{hi} + k_{hj} - k_{ij}) / 2 \sqrt{k_{hi} k_{hj}}$$

and the degree of asymmetry involved depends on these.

Similar considerations lead to the conclusion that the higher order levels of the proposed complete test and its power functions are reasonably close to the desired levels and functions existing in corresponding tests of uncorrelated and homoscedastic means and that the deviations involved appear to be on the conservative side.

REFERENCES

- [1] Bleicher, Edwin, "The Extension of the Multiple Comparisons Test to Lattice Designs," Virginia Polytechnic Institute M.S. thesis, 55 pp., 1954.
- [2] Cochran, W. G. and Cox, G. M., "Experimental Designs," 2nd Ed., John Wiley and Sons, Inc., New York, 1956.
- [3] Duncan, David B., "Multiple Range and Multiple F Tests," Biometrics, 11, 1-42, 1955.
- [4] Federer, Walter T., "Experimental Design," Macmillan, New York, 1955.
- [5] Keuls, M., "The Use of the 'Studentized Range' in Connection with an Analysis of Variance," Euphytica, 1, 112-122, 1952.
- [6] Kramer, Clyde Young, "Extension of Multiple Range Tests to Group Means with Unequal Numbers of Replications," Biometrics, 12, 307-310, 1956.
- [7] Kramer, Clyde Young, "Extension of Multiple Range Tests to Group Correlated Adjusted Means," Biometrics, 13, , 1957.
- [8] Newman, D., "The Distribution of the Range in Samples from a Normal Population, Expressed in Terms of an Independent Estimate of Standard Deviation," Biometrika, 31, 20-30, 1939.
- [9] Sanders, Paul G., "The Extension of the Multiple Comparisons Test to Rectangular Lattices," Virginia Polytechnic Institute M.S. thesis, 31 pp., 1953.
- [10] Tukey, J. W., "The Problem of Multiple Comparisons," unpublished dittoed notes, Princeton University, 396 pp., 1953.