A MODIFICATION OF THE NEWMAN-KEULS PROCEDURE FOR MULTIPLE COMPARISONS*

by

Roy E. Welsch

Working Paper 614-72

September 1972

*This research was supported in part by the Air Force/Rome Air Development Center through the Cambridge Project and by the National Bureau of Economic Research Computer Center for Economics and Management Science.

MOS Classification Numbers (1970):

Primary 62J15

Secondary 62Q05

Key Words: multiple comparisons, multiple-range tests, simultaneous statistical inference.

Abstract

A major criticism of the Newman-Keuls multiple comparison procedure is that it fails to provide adequate protection against erroneous comparisons when the null hypothesis of equal mean values is violated. This paper presents a modified Newman-Keuls procedure which ameliorates the above problem without, in the opinion of the author, becoming unduly conservative. Tables are provided which make the new test easy to use.

1. Introduction

A major criticism of the Newman-Keuls (N-K) multiple comparison procedure is that it fails to provide adequate protection against erroneous comparisons when the null hypothesis of equal mean values is violated. This was noticed by Tukey (1953) and Duncan (1955), who cites the following example. Suppose that in a 5% level N-K test of four means with $\nu = \infty$ and unit standard error, the values of the true means are $\mu_1 = \mu_2 = \mu$ and $\mu_3 = \mu_4 = \mu + \delta$. Let $m_1 \leq m_2 \leq m_3 \leq m_4$ be the sample mean values (m_1 does not necessarily come from the population with mean μ_1). Suppose that the difference δ between the two groups of means is so large that the preliminary range tests are practically certain to be significant, then the probability of jointly deciding that both $m_2 - m_1$ and $m_4 - m_3$ are not significant is $P\{m_2-m_1 \leq 2.77\} \cdot P\{m_4-m_3 \leq 2.77\} = 90.25\%$. The N-K test is described in detail in Miller (1966).

This paper presents a modified Newman-Keuls procedure which ameliorates the above problem without, in the opinion of the author, becoming unduly conservative. Tables are provided which make the new test easy to use.

2. Error Rates

We view multiple comparison procedures as mainly appropriate for the exploration of data rather than for decision-making. The test results are to be used to guide our thinking in the context of the problem at hand.

Generally we want to know if the data indicates that one population under consideration is better or worse than another. We must, therefore, compute some numbers to use as measures of significance (critical values). And this means making decisions about errors and error rates. Kurtz (1965), Miller (1966), and O'Neill (1971) discuss the question of error rates extensively.

For those interested in exploring data, a particular definition of error and error rate provides a way to compute a set of critical values and perhaps make some power calculations. It is certainly conceivable that a data analyst might use more than one set of critical values in analyzing a particular batch of data, weighing the results in light of the definition of error and error rate used to determine each set of critical values. Feder (1972) presents some new ways to make this easier in practice.

The author has generally not used the N-K test because the problem mentioned in the introduction makes it difficult to interpret the results in the light of any reasonable definition of error and error rate, especially when dealing with a fairly large number of means.

In contrast, the Tukey range test is based on the experimentwise error rate which is defined as the number of experiments with one or more erroneous comparisons divided by the number of experiments and,

as a consequence, provides a useful framework in which to view the results. An erroneous comparison occurs if two means are declared to be different when they are, in fact, equal or if the order of two unequal means is reversed.

Our test is also based on the experimentwise error rate but the definition of erroneous comparison is different. For us an erroneous comparison occurs only if two means are declared different when they are in fact equal. Thus in the determination of the error rate we do not count errors due to reversing the order of two unequal means.

The primary motivation for this approach is that generally in follow-on experiments such a reversal of order will be detected, but we do not want to spend money investigating differences that are not really there. In doing so we are, of course, making implicit assumptions about the relative costs of different errors. This should be taken into account is using the new test.

Given this definition of error we can expect that our test will be less stringent than the Tukey test but quite a bit more conservative than the N-K procedure.

3. Theory

Our problem can be stated formally as follows. The variates $M_1, M_2, \ldots, M_r$ and S are mutually independent, where M_i is $N(\mu_i, \sigma^2)$ and $\nu S^2/\sigma^2$ is χ^2_ν. There are unknown parameters $\lambda_1, \lambda_2, \ldots, \lambda_{r-1}$ and μ such that $\mu_1 = \mu$, $\mu_2 = \mu + \lambda_1$, $\mu_3 = \mu + \lambda_1 + \lambda_2, \ldots, \mu_r = \mu + \sum_{i=1}^{r-1} \lambda_i$. Generally μ is not of direct interest. We seek to determine if $\lambda_i = 0$ or $\lambda_i > 0$. The λ_i that are greater than zero can be viewed as determining blocks that we would like to separate. In reality we are probably only interested in finding which λ_i are larger than some multiple of σ. Since in most experiments with moderate r we might expect at least one large λ_i, we are interested in protecting ourselves from false positives about the λ_j $(j \neq i)$ due to the definite separation into blocks caused by a large value of λ_i (i.e. the problem mentioned earlier about the N-K test).

For r = 5 the possible hypotheses ($\lambda_i = 0$ vs. $\lambda_i > 0$) are (except for rearrangements):

(a) XXXXX

(b) XXXX X

(c) XXX XX

(d) XXX X X

(e) XX XX X

(f) XX X X X

(g) X X X X X

Assume that (c) holds so we actually have

H: $\mu_1 = \mu_2 = \mu_3 = \mu$ and $\mu_4 = \mu_5 = \mu + \lambda_1$

with $\lambda_1 > 0$.

Assume also that we want a series of critical numbers (depending on ν and r), $B_5 \geq B_4 \geq B_3 \geq B_2$, that we are going to use, just as in the N-K procedure to test groups of 5 means, 4 means, etc. in sequence. We also want to follow the rule that when a group of means has been declared not significantly different, we do not test any further subgroups. Let R_i be the range of i normal variates with equal means and variance σ^2, then we claim that

$$P\{\text{erroneous comparison}|H\} \leq P\{R_3 > SB_3\} + P\{R_2 > SB_2\}.$$

Let $m_1 \leq m_2 \leq \cdots \leq m_5$ and s be the sample values (recall m_1 does not necessarily come from the population with mean μ_1). Then comparing (1,5) with B_5 (i.e. m_5-m_1 with SB_5) we can only make an error if $R_3 > SB_5$ or $R_2 > SB_5$. Since $B_5 \geq B_3, B_2$ we have counted this error. Next we would look at (1,4) and (2,5). Again we only get an error if $R_3 > SB_4$, or $R_2 > SB_4$. For comparing (1,3), (2,4), (3,5) we use B_3 with error if $R_3 > SB_3$ or $R_2 > SB_3$. If (1,3) are declared significantly different we will look at (1,2) and (2,3). If m_1 and m_3 are from populations 1, 2, and 3, declaring (1,3) significant is an error and we have already counted this realization in our error rate. If (1,3) are not significantly different we do not go on to test (1,2) and (2,3). Similarly for (2,4) and (3,5). Hence we never use the value B_2 in checking groups of two means from the populations 1, 2, or 3 unless an error has been made at a previous stage. (It is crucial that the B_i's form an ordered sequence.)

Thus in order to control the error rate over experiments for r = 5 we want to choose $B_5 \geq B_4 \geq B_3 \geq B_2$ so that

$$P\{R_i \geq SB_i\} \leq .05 \qquad i=2,\ldots,5$$

and

$$2P\{R_2 > SB_2\} \leq .05$$

$$P\{R_3 > SB_3\} + P\{R_2 > SB_2\} \leq .05.$$

This accounts for all of the hypotheses (a) through (g). In order to achieve maximum power we would like the B_i's as small as possible.

Assuming we had an appropriate expression for the power and an analytic form of the cumulative distribution of the studentized range we might be able to solve for the B_i. We have neither. The case $r = 5$ discussed here reduces to determining p_2 and p_3 ($p_i = P\{R_i > SB_i\}$) so that

$$p_5 \leq .05, \; p_4 \leq .05$$

$$p_2 + p_3 \leq .05$$

$$2p_2 \leq .05$$

and

$$B_5 \geq B_4 \geq B_3 \geq B_2.$$

After considerable experimentation and some Monte Carlo studies of power, a reasonable procedure seemed to be to assign to each group of means a weight proportional to the number of means in the group. Thus for $r = 5$ we set $p_2 = (2/5)\cdot(.05)$, $p_3 = (3/5)\cdot(.05)$, but $p_4 = .05$ and $p_5 = .05$. In all cases used for tabling the values this led to $B_3 > B_2$. In a few cases we had $B_3 > B_4$ in which case we set $B_4 = B_3$.

The arguments used above apply to any r. For $r = 10$ the inequalities constraining the p_i are

$$p_i \le .05 \qquad i=2,3,\ldots,10$$

$$p_8 + p_2 \le .05$$

$$p_7 + p_3 \le .05,\ p_7 + p_2 \le .05$$

$$p_6 + p_4 \le .05,\ p_6 + 2p_2 \le .05,\ p_6 + p_3 \le .05$$

$$2p_5 \le .05,\ p_5 + p_3 + p_2 \le .05,\ p_5 + p_4 \le .05$$

$$2p_4 + p_2 \le .05,\ p_4 + 2p_3 \le .05,\ p_4 + 3p_2 \le .05,\ p_4 + p_3 + p_2 \le .05$$

$$3p_3 \le .05,\ 2p_3 + 2p_2 \le .05,\ p_3 + 3p_2 \le .05$$

$$5p_2 \le .05$$

$$B_i \ge B_{i-1} \qquad i=3,4,\ldots,10.$$

In this case we used $p_i = (i/10)\cdot(.05)$ except that p_{r-1} was set equal to .05. Table I lists the values of p_i for each r. Using these values, the only cases where $B_{i-1} > B_i$ occurred was with $i = r-1$ and in these cases B_{r-1} was set equal to B_{r-2}.

We used the method of inverse interpolation described in Harter (1959) to obtain the critical values from tables of the studentized range. The tables in this paper were computed to an accuracy of one unit in the fourth significant digit and then rounded to three significant digits. Linear harmonic ν-wise interpolation is recommended.

4. Examples

The tables that form a part of this paper can be used for the Tukey test, the N-K test, and the test described above. Consider the following data:

sample means:

8.29	8.96	12.57	15.41	16.50	18.92
m_1	m_2	m_3	m_4	m_5	m_6

standard error of m_i: 1.0

degrees of freedom: 30

An underscore is used to indicate that the means cannot be declared significantly different.

Tukey Test

k	Test Value	Comparisons
6	4.30	(1,6)
5	4.30	(1,5) (2,6)
4	4.30	(1,4) (2,5) (3,6)
3	4.30	(1,3) (2,4) (3,5) (4,6)
2	4.30	

The test value is just the single entry for k = 6, r = 6 in the table with $\upsilon = 30$.

N-K Test

k	Test Value	Comparisons
6	4.30	(1,6)
5	4.10	(1,5) (2,6)
4	3.85	(1,4) (2,5) (3,6)
3	3.49	(1,3) (2,4) (3,5) (4,6)
2	2.89	(1,2) (2,3) (3,4) (4,5) (5,6)

In this case the test values are the diagonal entries (k=6, r=6), (k=5, r=5),...,(k=2, r=2) in the same table.

New Test

k	Test Value	Comparisons
6	4.30	(1,6)
5	4.10	(1,5) (2,6)
4	4.10	(1,4) (2,5) (3,6)
3	3.92	(1,3) (2,4) (3,5) (4,6)
2	3.59	(1,2) (2,3) (3,4)

The test values are the entries in the r = 6 column of the same table.

Table I

Probabilities Used to Compute Critical Values

Size of Group (k)	Total Number of Means (r) 10	9	8	7	6	5	4	3	2
10	.0500								
9	.0500*	.0500							
8	.0400	.0500*	.0500						
7	.0350	.0389	.0500*	.0500					
6	.0300	.0333	.0375	.0500*	.0500				
5	.0250	.0278	.0312	.0357	.0500*	.0500			
4	.0200	.0222	.0250	.0286	.0333	.0500*	.0500		
3	.0150	.0167	.0187	.0214	.0250	.0300	.0500	.0500	
2	.0100	.0111	.0125	.0143	.0167	.0200	.0250	.0500	.0500

*Actual table entries in some cases are larger than this probability requires in order to insure that $B_{r-2} < B_{r-1}$.

Table II

DF= 5

Total Number of Means (r)

Size of Group (k)	10	9	8	7	6	5	4	3	2
10	6.99								
9	6.97	6.80							
8	6.97	6.75	6.58						
7	6.93	6.75	6.50	6.33					
6	6.87	6.69	6.50	6.19	6.03				
5	6.78	6.60	6.41	6.19	5.81	5.67			
4	6.61	6.44	6.26	6.05	5.81	5.30	5.22		
3	6.32	6.16	5.98	5.78	5.56	5.30	4.60	4.60	
2	5.70	5.55	5.39	5.21	5.00	4.76	4.47	3.64	3.64

DF= 6

Total Number of Means (r)

Size of Group (k)	10	9	8	7	6	5	4	3	2
10	6.49								
9	6.44	6.32							
8	6.44	6.25	6.12						
7	6.40	6.25	6.02	5.90					
6	6.33	6.18	6.02	5.74	5.63				
5	6.23	6.08	5.92	5.74	5.40	5.31			
4	6.07	5.93	5.77	5.60	5.40	4.94	4.90		
3	5.80	5.66	5.51	5.35	5.16	4.94	4.34	4.34	
2	5.24	5.12	4.98	4.83	4.65	4.44	4.20	3.46	3.46

DF= 7

Total Number of Means (r)

Size of Group (k)	10	9	8	7	6	5	4	3	2
10	6.16								
9	6.09	6.00							
8	6.09	5.91	5.82						
7	6.04	5.91	5.70	5.61					
6	5.97	5.84	5.70	5.45	5.36				
5	5.87	5.74	5.60	5.45	5.13	5.06			
4	5.71	5.59	5.46	5.30	5.13	4.70	4.68		
3	5.46	5.34	5.21	5.06	4.90	4.70	4.17	4.17	
2	4.95	4.84	4.72	4.58	4.42	4.24	4.02	3.34	3.34

DF= 8

Total Number of Means (r)

Size of Group (k)	10	9	8	7	6	5	4	3	2
10	5.92								
9	5.85	5.77							
8	5.85	5.68	5.60						
7	5.79	5.68	5.48	5.40					
6	5.72	5.60	5.48	5.24	5.17				
5	5.62	5.50	5.38	5.24	4.94	4.89			
4	5.46	5.36	5.23	5.10	4.94	4.54	4.53		
3	5.22	5.12	5.00	4.87	4.71	4.54	4.04	4.04	
2	4.75	4.65	4.53	4.41	4.27	4.10	3.89	3.26	3.26

DF= 9

Size of Group (k)	Total Number of Means (r): 10	9	8	7	6	5	4	3	2
10	5.74								
9	5.66	5.60							
8	5.66	5.50	5.43						
7	5.61	5.50	5.31	5.24					
6	5.53	5.43	5.31	5.08	5.02				
5	5.43	5.33	5.21	5.08	4.80	4.76			
4	5.28	5.18	5.07	4.94	4.80	4.41	4.41		
3	5.05	4.95	4.84	4.72	4.58	4.41	3.95	3.95	
2	4.60	4.50	4.40	4.28	4.15	3.99	3.80	3.20	3.20

DF= 10

Size of Group (k)	Total Number of Means (r): 10	9	8	7	6	5	4	3	2
10	5.60								
9	5.52	5.46							
8	5.52	5.36	5.30						
7	5.46	5.36	5.18	5.12					
6	5.39	5.29	5.18	4.96	4.91				
5	5.29	5.19	5.08	4.96	4.69	4.65			
4	5.14	5.05	4.94	4.82	4.69	4.33	4.33		
3	4.92	4.82	4.72	4.61	4.47	4.32	3.88	3.88	
2	4.48	4.39	4.30	4.19	4.06	3.91	3.73	3.15	3.15

DF= 11

Size of Group (k)	Total Number of Means (r) 10	9	8	7	6	5	4	3	2
10	5.49								
9	5.40	5.35							
8	5.40	5.25	5.20						
7	5.35	5.25	5.08	5.03					
6	5.27	5.18	5.08	4.86	4.82				
5	5.17	5.08	4.98	4.86	4.60	4.57			
4	5.03	4.94	4.84	4.73	4.60	4.26	4.26		
3	4.81	4.72	4.63	4.52	4.39	4.24	3.82	3.82	
2	4.39	4.31	4.22	4.11	3.99	3.84	3.67	3.11	3.11

DF= 12

Size of Group (k)	Total Number of Means (r) 10	9	8	7	6	5	4	3	2
10	5.40								
9	5.31	5.27							
8	5.31	5.16	5.12						
7	5.25	5.16	4.99	4.95					
6	5.18	5.09	4.99	4.79	4.75				
5	5.08	4.99	4.90	4.79	4.53	4.51			
4	4.94	4.86	4.76	4.65	4.53	4.20	4.20		
3	4.73	4.64	4.55	4.45	4.32	4.18	3.77	3.77	
2	4.32	4.24	4.15	4.05	3.93	3.79	3.62	3.08	3.08

DF= 13

Size of Group (k)	10	9	8	7	6	5	4	3	2
10	5.32								
9	5.23	5.19							
8	5.23	5.09	5.05						
7	5.18	5.09	4.92	4.88					
6	5.10	5.02	4.92	4.72	4.69				
5	5.00	4.92	4.83	4.72	4.47	4.45			
4	4.87	4.79	4.69	4.59	4.47	4.15	4.15		
3	4.66	4.58	4.49	4.39	4.27	4.13	3.73	3.73	
2	4.26	4.18	4.10	4.00	3.88	3.75	3.58	3.06	3.06

Total Number of Means (r) heads the columns 10–2.

DF= 14

Size of Group (k)	10	9	8	7	6	5	4	3	2
10	5.25								
9	5.17	5.13							
8	5.17	5.03	4.99						
7	5.11	5.03	4.86	4.83					
6	5.04	4.96	4.86	4.67	4.64				
5	4.94	4.86	4.77	4.67	4.42	4.41			
4	4.80	4.73	4.64	4.54	4.42	4.11	4.11		
3	4.60	4.52	4.43	4.34	4.22	4.09	3.70	3.70	
2	4.21	4.14	4.05	3.95	3.84	3.71	3.55	3.03	3.03

Total Number of Means (r) heads the columns 10–2.

DF= 15

Size of Group (k)	Total Number of Means (r) 10	9	8	7	6	5	4	3	2
10	5.20								
9	5.11	5.08							
8	5.11	4.97	4.94						
7	5.05	4.97	4.81	4.78					
6	4.98	4.90	4.81	4.62	4.59				
5	4.89	4.81	4.72	4.62	4.38	4.37			
4	4.75	4.68	4.59	4.49	4.38	4.08	4.08		
3	4.55	4.47	4.39	4.29	4.18	4.05	3.67	3.67	
2	4.17	4.09	4.01	3.92	3.81	3.68	3.52	3.01	3.01

DF= 16

Size of Group (k)	Total Number of Means (r) 10	9	8	7	6	5	4	3	2
10	5.15								
9	5.06	5.03							
8	5.06	4.93	4.90						
7	5.01	4.93	4.77	4.74					
6	4.93	4.86	4.77	4.58	4.56				
5	4.84	4.76	4.68	4.58	4.34	4.33			
4	4.71	4.63	4.55	4.45	4.34	4.05	4.05		
3	4.51	4.43	4.35	4.26	4.15	4.02	3.65	3.65	
2	4.13	4.06	3.98	3.89	3.78	3.65	3.50	3.00	3.00

DF= 17

Size of Group (k)	Total Number of Means (r) 10	9	8	7	6	5	4	3	2
10	5.11								
9	5.02	4.99							
8	5.02	4.89	4.86						
7	4.96	4.89	4.73	4.71					
6	4.89	4.82	4.73	4.55	4.52				
5	4.80	4.72	4.64	4.55	4.31	4.30			
4	4.67	4.59	4.51	4.42	4.31	4.02	4.02		
3	4.47	4.40	4.32	4.22	4.12	3.99	3.63	3.63	
2	4.10	4.03	3.95	3.86	3.76	3.63	3.48	2.98	2.98

DF= 18

Size of Group (k)	Total Number of Means (r) 10	9	8	7	6	5	4	3	2
10	5.07								
9	4.98	4.96							
8	4.98	4.85	4.82						
7	4.93	4.85	4.70	4.67					
6	4.85	4.78	4.70	4.51	4.49				
5	4.76	4.69	4.61	4.51	4.28	4.28			
4	4.63	4.56	4.48	4.39	4.28	4.00	4.00		
3	4.44	4.37	4.29	4.20	4.09	3.97	3.61	3.61	
2	4.07	4.00	3.92	3.84	3.73	3.61	3.46	2.97	2.97

DF= 19

Size of Group (k)	Total Number of Means (r) 10	9	8	7	6	5	4	3	2
10	5.04								
9	4.95	4.92							
8	4.95	4.82	4.79						
7	4.89	4.82	4.67	4.65					
6	4.82	4.75	4.67	4.49	4.47				
5	4.73	4.66	4.58	4.49	4.26	4.25			
4	4.60	4.53	4.45	4.36	4.26	3.98	3.98		
3	4.41	4.34	4.26	4.17	4.07	3.95	3.59	3.59	
2	4.05	3.98	3.90	3.81	3.71	3.59	3.44	2.96	2.96

DF= 20

Size of Group (k)	Total Number of Means (r) 10	9	8	7	6	5	4	3	2
10	5.01								
9	4.92	4.90							
8	4.92	4.79	4.77						
7	4.86	4.79	4.64	4.62					
6	4.79	4.72	4.64	4.46	4.45				
5	4.70	4.63	4.55	4.46	4.23	4.23			
4	4.57	4.50	4.43	4.34	4.23	3.96	3.96		
3	4.38	4.31	4.24	4.15	4.05	3.93	3.58	3.58	
2	4.02	3.96	3.88	3.80	3.70	3.58	3.43	2.95	2.95

DF= 24

Total Number of Means (r)

Size of Group (k)	10	9	8	7	6	5	4	3	2
10	4.92								
9	4.83	4.81							
8	4.83	4.70	4.68						
7	4.77	4.70	4.56	4.54					
6	4.70	4.63	4.56	4.38	4.37				
5	4.61	4.54	4.47	4.38	4.17	4.17			
4	4.49	4.42	4.35	4.26	4.17	3.90	3.90		
3	4.30	4.24	4.16	4.08	3.98	3.87	3.53	3.53	
2	3.96	3.89	3.82	3.74	3.64	3.52	3.38	2.92	2.92

DF= 30

Total Number of Means (r)

Size of Group (k)	10	9	8	7	6	5	4	3	2
10	4.82								
9	4.74	4.72							
8	4.74	4.62	4.60						
7	4.68	4.62	4.48	4.46					
6	4.61	4.55	4.48	4.31	4.30				
5	4.52	4.46	4.39	4.31	4.10	4.10			
4	4.40	4.34	4.27	4.19	4.10	3.85	3.85		
3	4.22	4.16	4.09	4.01	3.92	3.81	3.49	3.49	
2	3.89	3.83	3.76	3.68	3.59	3.48	3.34	2.89	2.89

DF= 40

Total Number of Means (r)

Size of Group (k)	10	9	8	7	6	5	4	3	2
10	4.73								
9	4.65	4.63							
8	4.65	4.53	4.52						
7	4.59	4.53	4.40	4.39					
6	4.53	4.47	4.40	4.24	4.23				
5	4.44	4.38	4.31	4.24	4.04	4.04			
4	4.32	4.26	4.20	4.12	4.03	3.79	3.79		
3	4.15	4.09	4.02	3.95	3.86	3.75	3.44	3.44	
2	3.82	3.77	3.70	3.62	3.53	3.43	3.29	2.86	2.86

DF= 60

Total Number of Means (r)

Size of Group (k)	10	9	8	7	6	5	4	3	2
10	4.65								
9	4.56	4.55							
8	4.56	4.45	4.44						
7	4.51	4.45	4.32	4.31					
6	4.44	4.38	4.32	4.17	4.16				
5	4.36	4.30	4.24	4.17	3.98	3.98			
4	4.24	4.19	4.12	4.05	3.97	3.74	3.74		
3	4.07	4.02	3.95	3.88	3.80	3.70	3.40	3.40	
2	3.76	3.71	3.64	3.57	3.48	3.38	3.25	2.83	2.83

DF= 120

Size of Group (k)	Total Number of Means (r): 10	9	8	7	6	5	4	3	2
10	4.56								
9	4.48	4.47							
8	4.48	4.37	4.36						
7	4.42	4.37	4.25	4.24					
6	4.36	4.31	4.25	4.10	4.10				
5	4.28	4.22	4.16	4.10	3.92	3.92			
4	4.17	4.11	4.05	3.98	3.90	3.68	3.68		
3	4.00	3.95	3.89	3.82	3.74	3.64	3.36	3.36	
2	3.70	3.65	3.59	3.52	3.43	3.33	3.21	2.80	2.80

DF= ∞

Size of Group (k)	Total Number of Means (r): 10	9	8	7	6	5	4	3	2
10	4.47								
9	4.39	4.39							
8	4.39	4.29	4.29						
7	4.34	4.29	4.17	4.17					
6	4.28	4.23	4.17	4.03	4.03				
5	4.20	4.15	4.09	4.03	3.86	3.86			
4	4.09	4.04	3.98	3.92	3.84	3.63	3.63		
3	3.93	3.88	3.83	3.76	3.68	3.59	3.31	3.31	
2	3.64	3.59	3.53	3.46	3.39	3.29	3.17	2.77	2.77

References

Duncan, D. B. (1955). Multiple range and multiple F tests. Biometrics, 11, 1-42.

Feder, P. I. (1972). Studentized range graph paper--a new tool for the graphical comparison of treatment means. TIS Report No. 72-CRD-193. General Electric Co., Schenectady, N.Y.

Harter, H. L., Clemm, D. S., and Guthrie, E. H. (1959). The probability integrals of the range and of the studentized range-probability integral and percentage points of the studentized range; critical values for Duncan's new multiple range test. Wright Air Development Center Technical Report 58-484, Vol. II. (ASTIA Document No. AD231733).

Kurtz, T. E., Link, R. F., Tukey, J. W. and Wallace, D. L. (1965). Short-cut multiple comparisons for balanced single and double classifications: Part 1, Results. Technometrics, 7, 95-165.

Miller, R. G. (1966). Simultaneous Statistical Inference. New York: McGraw-Hill.

O'Neill, R. and Wetherill, G. B. (1971). The present state of multiple comparison methods. JRSS B, 33, 218-250.

Tukey, J. W. (1953). The problem of multiple comparisons. Unpublished Dittoed Notes, Princeton University.