TABLE OF COEFFICIENTS OF ASSOCIATION

.7368	.7894	.6842	.6315	.7368	.8157
.6315	.7894	.7105	.6842	.8157	.7894
.7631	.6842	.8684	.5789	.7368	.7368
.6842	.6842	.7631	.7105	.7631	.6842
.7894	.8421	.8421	.8421	.8684	.8157
.7894	.8157	.7894	.8157	.7631	.8157

.7368	.7368	.7368	.7894	.7105	.6578
.6842	.6052	.6842	.6578	.6315	.6052
.6842	.7105	.5263	.7894	.7368	.7631
.6842	.7105	.6578	.6578	.6315	.6578
.7894	.7368	.7368	.7631	.7631	.7368
.7631	.7368	.7631	.7105	.7631	.8157

.6052	.5526	.7631	.8947	.6842	.6052
.7368	.6578	.8421	.7105	.5263	.7105
.8421	.5526	.7105	.7105	.9473	.8947
.8947	.8421	.7368	.8157	.8421	.8684
.9210	.9210	.9473	.9473	.9210	.9210
.9210	.8947	.8947	.9473	.8947	.8684

.6578	.7631	.8421	.6842	.6052	.8421
.7105	.7894	.7631	.5789	.7105	.6578
.6052	.7105	.7105	.8421	.7894	.7894
.7368	.7894	.7105	.7368	.7631	.7631
.8157	.8421	.8421	.8157	.8157	.8421
.8421	.7894	.8421	.7894	.7631	

.6842	.6578	.7631	.7368	.6578	.7105
.6578	.6315	.6052	.6842	.5789	.6315
.7368	.6842	.7105	.7105	.6578	.6842
.6578	.5789	.6052	.6315	.6315	.6842
.7105	.7105	.6842	.6842	.7105	.7368
.7105	.7105	.8157	.6315		

.7631	.5526	.5263	.7631	.7105	.6315
.8315	.5000	.5263	.5789	.5263	.6052
.8315	.8157	.7631	.8157	.8421	.7894
.7894	.8157	.8421	.7894	.7894	.7894
.8157	.7894	.7894	.8684	.7894	.7894
.8157	.8157	.7894			

.7631	.7894	.6052	.6578	.6315	.6315
.8052	.7368	.6315	.7368	.6052	.6842
.8526	.5000	.5526	.4736	.5263	.5789
.8473	.4736	.5263	.5789	.5789	.6315
.8263	.5789	.5526	.5789	.4736	.6052
.8052	.5789				

ROW NUMBER = 1

1.0000	.8421	.8157	.9210	.7368	.7894	.6842	.6315	.7368	.8157
.7105	.5789	.8157	.7631	.6315	.7894	.7105	.6842	.8157	.7894
.6052	.7368	.6842	.6315	.7631	.6842	.8684	.5789	.7368	.7368
.8157	.7631	.7631	.7368	.6842	.6842	.7631	.7105	.7631	.6842
.7105	.7368	.7368	.8421	.7894	.8421	.8421	.8421	.8684	.8157
.7894	.7894	.8157	.8421	.7894	.8157	.7894	.8157	.7631	.8157
.8157	.7894								

ROW NUMBER = 2

1.0000	.8157	.8684	.7894	.7368	.7368	.7368	.7894	.7105	.6578
.6315	.7631	.7631	.5263	.6842	.6052	.6842	.6578	.6315	.6052
.6842	.5789	.6315	.6578	.6842	.7105	.5263	.7894	.7368	.7631
.71 5	.7631	.6842	.6315	.6842	.7105	.6578	.6578	.6315	.6578
.6842	.6842	.7894	.7368	.7894	.7368	.7368	.7631	.7631	.7368
.7894	.7631	.7894	.6842	.7631	.7368	.7631	.7105	.7631	.8157
.6842									

ROW NUMBER = 3

1.0000	.8421	.7631	.8157	.6052	.5526	.7631	.8947	.6842	.6052
.8421	.7894	.6052	.8157	.7368	.6578	.8421	.7105	.5263	.7105
.7631	.6578	.7894	.7105	.8421	.5526	.7105	.7105	.9473	.8947
.8421	.8684	.8157	.8684	.8947	.8421	.7368	.8157	.8421	.8684
.8684	.9210	.9210	.9210	.9210	.9210	.9473	.9473	.9210	.9210
.9473	.9210	.8684	.9473	.9210	.8947	.8947	.9473	.8947	.8684

ROW NUMBER = 4

1.0000	.7105	.7631	.6052	.6578	.7631	.8421	.6842	.6052	.8421
.7894	.6578	.7631	.6842	.7105	.7894	.7631	.5789	.7105	.6578
.6052	.7368	.6578	.8421	.6052	.7105	.7105	.8421	.7894	.7894
.7631	.7105	.7105	.7894	.7368	.7894	.7105	.7368	.7631	.7631
.8684	.8157	.8157	.8684	.8157	.8421	.8421	.8157	.8157	.8421
.8684	.8157	.8421	.8157	.8421	.7894	.8421	.7894	.7631	

ROW NUMBER = 5

1.0000	.6315	.7894	.6842	.6842	.6578	.7631	.7368	.6578	.7105
.5789	.5789	.5526	.5789	.6578	.6315	.6052	.6842	.5789	.6315
.6578	.7368	.6052	.5789	.7368	.6842	.7105	.7105	.6578	.6842
.6315	.7368	.6578	.6578	.6578	.5789	.6052	.6315	.6315	.6842
.6842	.7894	.7368	.7368	.7105	.7105	.6842	.6842	.7105	.7368
.6315	.7631	.6842	.7105	.7105	.7105	.8157	.6315		

ROW NUMBER = 6

1.0000	.5263	.5263	.6315	.7631	.5526	.5263	.7631	.7105	.6315
.7368	.6578	.5263	.7105	.6315	.5000	.5263	.5789	.5263	.6052
.6315	.7631	.6315	.6315	.6315	.8157	.7631	.8157	.8421	.7894
.7368	.7631	.6315	.7631	.7894	.8157	.8421	.7894	.7894	.7894
.8421	.7894	.7105	.7894	.8157	.7894	.7894	.8684	.7894	.7894
.8157	.7894	.7631	.7631	.8157	.8157	.7894			

ROW NUMBER = 7

1.0000	.7368	.7894	.5000	.7631	.7894	.6052	.6578	.6315	.6315
.7105	.6842	.6052	.7368	.6052	.7368	.6315	.7368	.6052	.6842
.6578	.5789	.7368	.6842	.5526	.5000	.5526	.4736	.5263	.5789
.5526	.5000	.5000	.4210	.4473	.4736	.5263	.5789	.5789	.6315
.5263	.5789	.6052	.5526	.5263	.5789	.5526	.5789	.4736	.6052
.5263	.5526	.5526	.5526	.6052	.5789				

Fortran II Program for Coefficient of Association (Match-Coeff) Using an IBM 1620 Computer

By
Roger L. Kaesler
Floyd W. Preston
and
Donald Good

SPECIAL DISTRIBUTION PUBLICATION 4

State Geological Survey
The University of Kansas, Lawrence, Kansas
1963

STATE GEOLOGICAL SURVEY OF KANSAS

W. Clarke Wescoe, M.D., Chancellor of The University and ex officio Director of the Survey

Frank C. Foley, Ph.D., State Geologist and Director

William W. Hambleton, Ph.D., Assoc. State Geologist and Assoc. Director

Raymond C. Moore, Ph.D., Sc.D., Principal Geologist Emeritus

John M. Jewett, Ph.D., Senior Geologist

Doris E. Nodine Zeller, Ph.D., Geologist, Editor

Grace M. Muilenburg, B.S., Public Information Director

Kenneth J. Badger, Chief Draftsman

Lila M. Watkins, Secretary

Research Divisions

Basic Geology.....Daniel F. Merriam, Ph.D., geologist in charge

Petrography and Geochemistry.....Ada Swineford, Ph.D., petrographer in charge

Mineral Resources.....Allison L. Hornbaker, M.S., geologist in charge

Oil and Gas..... Edwin D. Goebel, M.S., geologist in charge

Ceramics.....Norman Plummer, A.B., ceramist in charge

Cooperative Studies with United States Geological Survey

Ground-Water Resources.....Robert J. Dingman, B.S., geologist in charge

Mineral Fuels..... William D. Johnson, Jr., B.S., geologist in charge

Branch Offices

Geological Survey Well Sample Library, 4150 Monroe, Wichita

Geological Survey Southeast Kansas District Field Office, College Station, Pittsburg

Geological Survey Southwest Kansas Field Office, 310 N. 9th Street, Garden City

INTRODUCTION

MATCH-COEFF is a program for computing the coefficient of association or simple matching coefficient for as many as 70 items (samples, wells, operational taxonomic units, etc.) with as many as 70 characters or observations in each. The coefficient of association is a means of comparing samples in which the data can be reduced to 2 categories, such as present or absent, yes or no, etc. (Sokal and Michener, 1958, p. 1417). A number of characters is examined for each sample (preferably 35 or more) and a 2 is recorded for the property or condition if present, a 1 for the property or condition if absent. Each column of data (sample) is then compared with all others, and the number of character states each column has in common with every other column is recorded. Character states in common are called "matches." If data from two samples from a larger study were

2	1
1	2
1	1
2	2
2	2

the number of matches would be 3. This number of matches is divided by the total number of comparisons (5 in this case) to give the coefficient of association or simple matching coefficient (here 3/5 = 0.6).

When comparing more than 2 samples, one should not use data that show no change of state in a row (Sokal and Sneath, in preparation). The reason for this restriction is obvious if we consider an extreme hypothetical case.

Suppose that Pennsylvanian cyclothems have been sampled for the following fossils - Olenellus (Cambrian), Strophomena (Ordovician), Polydiexodina (Permian), and Acanthoscaphites (Cretaceous). Because none of these fossils were found, the data would have perfect matches at all localities, and the computed coefficients of association would equal one.

Similarly, in a mineralogical study of sandstone it generally would be unwise to include quartz because it would normally be present in all samples and would lead to an unjustifiably large coefficient of association.

Kaesler's work on this project was supported in part by a National Science Foundation Summer Fellowship for graduate teaching assistants. Computer operation expenses were defrayed by an allocation from the Computation Center, The University of Kansas.

DESCRIPTION OF MATCH-COEFF

The coefficient of association or simple matching coefficient indicates degree of similarity among samples with data reduced to two states (Sokal and Michener, 1958).

The equation for the coefficient is

$$S_{sm} = \frac{m}{n}$$

where S_{sm} = the coefficient
m = the number of matches
n = the total number of comparisons.

Nomenclature used in the program is shown in Table 1, Figure 1 is a flow chart of the program, and a listing of Fortran II statements is given in Table 2.

For a limited time, the Kansas Survey will make available the program deck of punched cards for the price of $2.00.

Table 1.--Listing of nomenclature.

M	=	number of columns.
N	=	number of rows.
I	=	row index.
J	=	column index.
K	=	dummy column index for comparison with J.
ISUM	=	number of matches minus number of mismatches.
S(K)	=	floating point equivalent of ISUM.
T(K)	=	coefficient of association or simple matching coefficient. K is incremented from J = 1 to M and is punched for each value of J from 1 to M. Diagonal values are set equal to 1.0000.
IZ(K)	=	input data found on one card, K = 1, 2, ..., M.
ID(J, K)	=	data matrix stored in core of computer; the 1's have been converted to -1 and the 2's to +1.

Table 2.--FORTRAN II statements in MATCH-COEFF program.

```
*0704
C      PREPARATION OF MATRIX OF COEFFICIENTS OF ASSOCIATION OR SIMPLE
C      MATCHING COEFFICIENTS
C      ROGER L. KAESLER, FLOYD W. PRESTON, AND DONALD I. GOOD
C      UNIVERSITY OF KANSAS
C      LAWRENCE, KANSAS
C
C      PROGRAM REQUIRES DATA AS IZ(I,K) WHERE ELEMENTS ARE 2 FOR PROPERTY
C      OR CONDITION PRESENT AND 1 FOR PROPERTY OR CONDITION ABSENT.
C      THESE 2S AND 1S ARE CONVERTED TO +1 AND -1 IN THE COMPUTER.
C
C      OUTPUT IS THE ASSOCIATION OR MATCHING COEFFICIENT MATRIX.  OUTPUT
C      IS NOT IN MATRIX FORM BECAUSE THE SIZE OF THE MATRIX MAY EXCEED
C      THE CAPACITY OF THE TABULATOR.  THE MATRIX IS A SYMMETRICAL M*M
C      MATRIX WITH M EQUAL TO THE NUMBER OF COLUMNS IN A Q-TYPE STUDY.
C      FOR NUMERICAL TAXONOMY WORK COLUMNS = OTUS, ROWS = CHARACTERS.
C      FOR ECOLOGICAL WORK, COLUMNS = STATIONS, ROWS = SPECIES.
C
C      M = NUMBER OF COLUMNS.
C      N = NUMBER OF ROWS.
C      I = ROW INDEX.
C      J = COLUMN INDEX.
C
```

```
      DIMENSION ID(70,70), IZ(70), S(70), T(70)
   10 READ 11,M,N
   11 FORMAT(10X,I2,2X,I2)
   12 AN= N
   13 ANI= 1./AN
   40 DO 140 I= 1,N,1
   60 READ 61, (IZ(K), K= 1,M,1)
   61 FORMAT(10X, 70I1)
   90 DO 130 J= 1,M,1
  100 IF (IZ(J)-1) 110,110,120
  110 ID(I,J)= -1
  115 GO TO 130
  120 ID(I,J)= 1
  130 CONTINUE
  140 CONTINUE
C
C     AT THIS POINT 1 AND -1 HAVE BEEN SUBSTITUTED FOR VALUES OF
C     2 AND 1 IN THE CORE.
C
  150 PUNCH 151
  151 FORMAT(22X,36HTABLE OF COEFFICIENTS OF ASSOCIATION//)
  160 DO 290 J= 1,M,1
  170 DO 270 K= 1,M,1
  180 ISUM= 0
  190 IF (J-K) 200, 260, 270
  200 DO 220 I= 1,N,1
  210 ISUM= ISUM + ID(I,J)*ID(I,K)
  220 CONTINUE
  230 S(K)= ISUM
  240 T(K)= 0.5*S(K)*ANI + 0.5
  250 GO TO 270
  260 T(K)= 1.0000
  270 CONTINUE
  280 PUNCH 281, J
  281 FORMAT(12X,12HROW NUMBER =, I3)
  285 PUNCH 286, (T(K), K= J,M,1)
  286 FORMAT(10X, 10F7.4)
  290 CONTINUE
  300 GO TO 10
      END
```

SAMPLE PROBLEM

In an area being investigated ecologically, samples were obtained at 5 stations. Six ostracode species were found in the samples, but not all were present at each station. Using 2 for "ostracode present" and 1 for "ostracode absent," the data are:

Station	1	2	3	4	5
Species					
1	2	1	2	1	2
2	2	1	1	2	2
3	2	1	2	2	1
4	2	1	1	1	1
5	2	1	2	1	2
6	1	2	1	2	2

Here M = 5, N = 6.

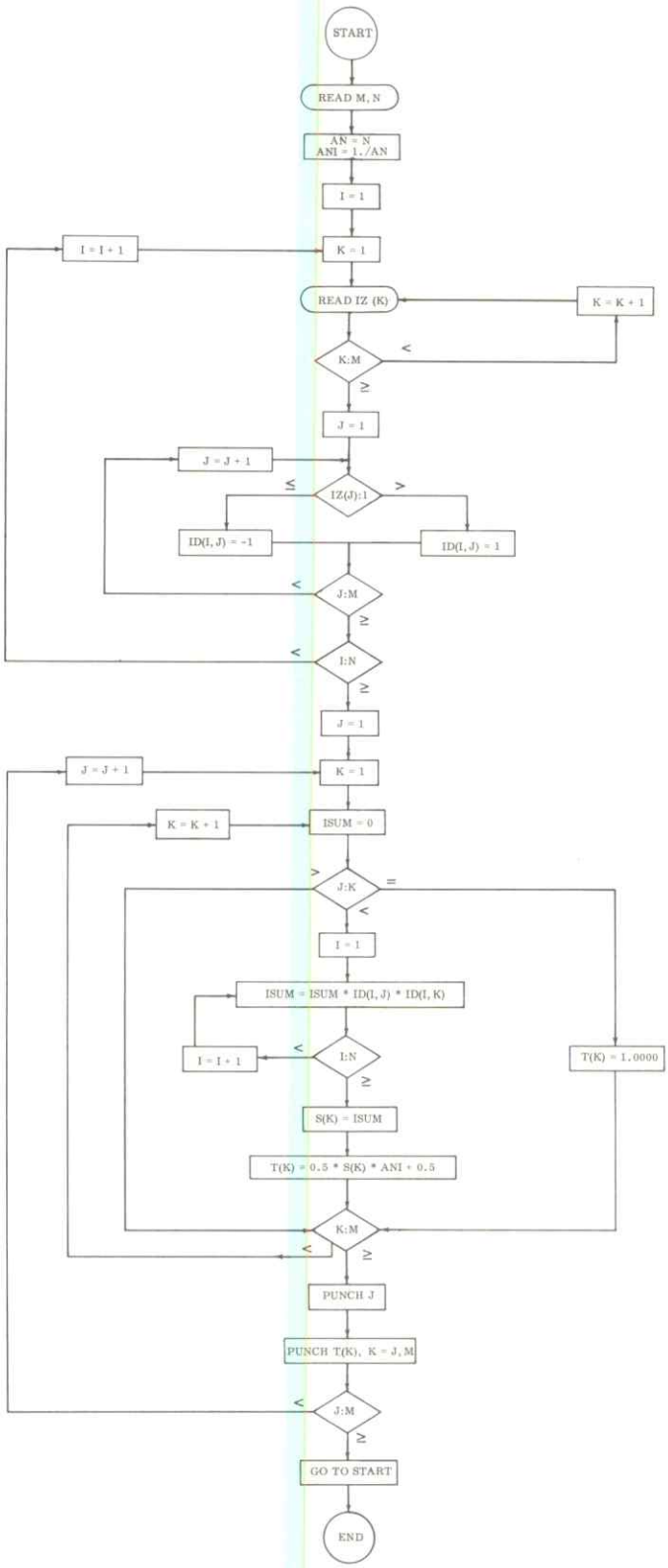

Figure 1. Flow chart of MATCH-COEFF program.

Using MATCH-COEFF to find coefficients of association, the input is:

```
0000        05  06
0001        21212
0002        21122
0003        21221
0004        21111
0005        21212
0006        12122
```

Output from the computations is:

```
                TABLE OF COEFFICIENTS OF ASSOCIATION

 ROW NUMBER =  1
1.0000 0.0000  .6666  .3333  .5000
 ROW NUMBER =  2
1.0000  .3333  .6666  .5000
 ROW NUMBER =  3
1.0000  .3333  .5000
 ROW NUMBER =  4
1.0000  .5000
 ROW NUMBER =  5
1.0000
```

Thus sample 1 is most closely related to sample 3 ($S_{sm} = 0.6666$); sample 2 is related to sample 4 to the same degree. Samples 1 and 2 have a coefficient of association of 0.0000 and are thus alike in none of the respects for which they are being tested, i.e., presence of the six ostracode species. Sample 5 has a coefficient of association of 0.5000 with every sample in the study.

NOTES FOR PROGRAM USER

1. The simple matching coefficient requires data reduced to 2 states. Examples of possible uses might be:
    a. species present or absent in a sample.
    b. heavy mineral present or absent.
    c. producing well or dry hole.
    d. limestone or sandstone.
2. The first card is a parameter card giving M and N.
3. To save space on the input card, 2 is used for a property or condition present, 1 for property or condition absent. These values are converted to 1 and -1 respectively by the computer.
4. Data must be of the form ID(I,J). For example, ID(3,5) is the 2 or 1 value for the third species of the fifth sample (the value of ID(3,5) is 2 in the sample problem).

5. Output is of two types. First, a card is punched telling what row of the matching coefficient matrix is to be punched next. This number will range from 1 to M. Second, cards with simple matching coefficients are punched with ten coefficients per card. The coefficient T(K) is punched with K increasing from J (the row number, where T(K) = 1.0000) to M. Only that part of the matrix which lies above the main diagonal is punched.

6. Because of the size of the matrix and the capacity of the cards, output is not in matrix form.

7. Interpretation of the simple matching coefficient matrix by itself is difficult. The output from this program is in proper format to be used as input to a clustering program; however, before the data can be clustered the J cards must be removed (see item 5 above).

8. Data which show no change of state along a row, that is, which contain all 1's or all 2's in one row, should not be used.

9. Decisions on similarity and differences among samples should be based on as many characters as possible---preferably 35 or more.

10. Cards for this program must be punched in the following format (the symbolism below refers to FORMAT statements of FORTRAN II):

    a. parameter card: 10X, I2, 2X, I2

    b. input data card: 10X, 70I1

    c. output coefficient card: 10X, 10F7.4

11. The program is written to accommodate a 70 by 70 data matrix. Any smaller size could be used; a larger size could be used only by changing the dimension statement and statement 61 and recompiling.

12. Computation of a 70 by 70 matrix takes about one hour.

REFERENCES

Sokal, R. R., and Michener, C. D., 1958, A statistical method for evaluating systematic relationships. Univ. of Kansas Science Bull., v. 38, pt. 2, p. 1409-1438.

Sokal, R. R., and Sneath, P. H. A., in preparation, The principles of numerical taxonomy.

KANSAS GEOLOGICAL SURVEY COMPUTER PROGRAM
THE UNIVERSITY OF KANSAS, LAWRENCE

PROGRAM ABSTRACT

Title (If subroutine state in title):

Coefficient of Association (Match-Coeff)

Computer: IBM 1620 Date: September 27, 1963

Programming language: Fortran II

Author, organization: Roger L. Kaesler, Department of Geology, The University of Kansas; Floyd W. Preston and Donald Good, Kansas Geological Survey

Direct inquiries to: Kansas Geological Survey, The University of Kansas

Name: Daniel F. Merriam Address: Kansas Geological Survey, The University of Kansas, Lawrence, Kansas

Purpose/description: Program compares 2 to 70 samples on which 2 to 70 variables have been recorded. Only two categories (e.g., +, -; or present, absent; ...) are permitted for each variable. The coefficient of association is calculated for each possible sample-to-sample comparison. This coefficient is the fraction of the variables in a sample pair which have identical "states" or values.

Mathematical method: Not applicable

Restrictions, range: Maximum of 70 samples and 70 variables. Number of samples and number of variables need not be equal.

Storage requirements: Program and data occupy approximately 70 percent of the memory space of a 60K (digit position) IBM 1620.

Equipment specifications:

Memory 20K____ 40K____ 60K x K____

Automatic divide: Yes x No____ Indirect addressing: Yes x No____

Other special features required____

Additional remarks (include at author's discretion: fixed/float, relocatability; optional: running time, approximate number of times run successfully, programming hours) A 70 sample x 36 variable data matrix requires about 1 hour of machine time.

Special
Distribution
Publication

1. The Kansas mineral industry . . . 1962; with directory of Kansas mineral producers, by Grace Mui R. G. Hardy, and Allison Hornbaker, 1963.
2. Economic development for Kansas, mineral and water resources: Report of the Governor's Economic Deve ment Committee, by W. W. Hambleton, and others, 1962.
3. BALGOL program for trend-surface mapping using an IBM 7090 computer, by J. W. Harbaugh, 1963.
4. FORTRAN II program for coefficient of association (Match-Coeff) using an IBM 1620 computer, by R. L. Kaesler, F. W. Preston, and Donald Good, 1963.

```
 ROW NUMBER =   1
1.0000   .8421   .8157   .92
 .7105   .5789   .8157   .76
 .6052   .7368   .6842   .63
 .8157   .7631   .7631   .73
 .7105   .7368   .7368   .84
 .7894   .7894   .8157   .84
 .8157   .7894
 ROW NUMBER =   2
1.0000   .8157   .8684   .78
 .6315   .7631   .7631   .52
 .6842   .5789   .6315   .65
 .7105   .7631   .6842   .63
 .6842   .6842   .7894   .73
 .7894   .7631   .7894   .68
 .6842
 ROW NUMBER =   3
1.0000   .8421   .7631   .81
 .8421   .7894   .6052   .81
 .7631   .6578   .7894   .71
 .8421   .8684   .8157   .86
 .8684   .9210   .9210   .92
 .9473   .9210   .8684   .94
 ROW NUMBER =   4
1.0000   .7105   .7631   .60
 .7894   .6578   .7631   .68
 .6052   .7368   .6578   .84
 .7631   .7105   .7105   .78
 .8684   .8157   .8157   .86
 .8684   .8157   .8421   .81
 ROW NUMBER =   5
1.0000   .6315   .7894   .68
 .5789   .5789   .5526   .57
 .6578   .7368   .6052   .57
 .6315   .7368   .6578   .65
 .6842   .7894   .7368   .73
 .6315   .7631   .6842   .71
 ROW NUMBER =   6
1.0000   .5263   .5263   .63
 .7368   .6578   .5263   .71
 .6315   .7631   .6315   .63
 .7368   .7631   .7105   .76
 .8421   .7894   .7894   .81
 .8157   .7894   .7631   .76
 ROW NUMBER =   7
1.0000   .7368   .7894   .50
 .7105   .6842   .6052   .73
 .6578   .5789   .7368   .68
 .5526   .5000   .5000   .42
 .5263   .5789   .6052   .55
 .5263   .5526   .5526   .55
```