A COMPUTER PROGRAM FOR THE ANALYSIS OF
TWO-WAY CONTINGENCY TABLES

by

Dana Quade
Department of Biostatistics
University of North Carolina School of Public Health

Institute of Statistics Mimeo Series No. 567

February 1968

NOTICE

This program has been placed in the Call-A-Computer library under the name ANATAB*** (for "ANAlysis of TABulation"). The former program TABLE* and its writeup are hereby superseded.

ERRATUM

The formula for $\hat{\theta}_i$ on page 10 is incorrect and should be replaced by

$$\hat{\theta}_i = \frac{2}{N} (\bar{X}_i - \frac{N+1}{2}) \quad .$$

Dana Quade
March 5, 1968

SUMMARY

This program, written in Time-sharing Fortran for Call-a-Computer, may be used for the analysis of two-way contingency tables with up to 50 rows, and up to 25 columns, and up to 524287 observations. Empty rows are not permitted. The program includes

1) Input of data
2) (Option) Printout of table
3) (Option) Expected values and "contributions to chi-square"
4) Chi-square
5) Fisher's test (for 2x2 tables)
6) (Option) Tests of symmetry (for square tables)
7) (Option) Comparison of rows
8) (Option) Correlation analysis
9) (Option) Repeat same analysis with different data, or reset.

A complete listing of the program is given as an Appendix.

DETAILED DESCRIPTION OF THE PROGRAM

Throughout this write-up the following notation is used: the number of rows is I, the number of columns J. The number of observations in the (i,j) cell is n_{ij}, $1 \leq i \leq I$, $1 \leq j \leq J$. The row totals are R_i, $1 \leq i \leq I$, the column totals are C_j, $1 \leq j \leq J$, and the grand total number of observations is N.

Consider the following contingency table as an example:

3	7	8	2	20
4	5	9	6	24
3	1	2	11	17
2	7	6	8	23
12	20	25	27	84

For this table I = J = 4 and N = 84.

1. INPUT OF DATA

The program asks first for the number of rows and columns in the table and then requests the data for each row, one at a time. Thus the result of entering the example table might appear as follows:

```
HOW MANY ROWS (2 TO 50) AND COLUMNS (2 TO 25)?
? 4 4

ENTER ROW 1
? 3 7 8 2

ENTER ROW 2
? 4 5 9 6

ENTER ROW 3
? 3 1 2 11

ENTER ROW 4
? 2 7 6 8
```

For small tables, such as the example, it is convenient to feed in all the data in response to the first question, since this avoids returning control to the computer for each row. Then the result would appear thus:

```
HOW MANY ROWS (2 TO 50) AND COLUMNS (2 TO 25) ?
? 4 4 3 7 8 2 4 5 9 6 3 1 2 11 2 7 6 8

ENTER ROW 1

ENTER ROW 2

ENTER ROW 3

ENTER ROW 4
```

Data entries must be separated by spaces, as shown, or (optionally) by commas. If the table to be analyzed is very large, or if there are a great number of tables to be done, it will be convenient to prepare a data tape in advance.

Note that the maximum capacity of the program is N = 524287 observations. A running total is kept as the data is fed in, and if this ever exceeds 524287 the program comes to an immediate halt, with the error message "STOP BIG". The program also halts if any attempt is made to enter a negative cell frequency (error message "STOP NEG") or an empty row (error message "STOP ROW"). Empty columns are permitted, however.

2. (OPTION) PRINTOUT OF TABLE

The example table will appear as follows if this option is taken:

DO YOU WANT THE TABLE PRINTED OUT?
?YES

R\C	1	2	3	4	TOTAL
1	3	7	8	2	20
2	4	5	9	6	24
3	3	1	2	11	17
4	2	7	6	8	23
TOTAL	12	20	25	27	84

If the table has more than 8 columns ($J > 8$) it will be "folded over" and "interlaced" and hence somewhat difficult to read. It will sometimes be convenient to avoid this, if the number of rows is 8 or fewer, by entering the data with rows and columns transposed.

Note: The grand total number of observations (N) is always given, whether or not the whole table is printed out.

3. (OPTION) EXPECTED VALUES AND "CONTRIBUTIONS TO CHI-SQUARE"

If this option is taken for the example table, the output will appear as follows:

DO YOU WANT EXPECTED VALUES?
?YES

ROW	COLUMN	OBSERVED	EXPECTED	(O-E)↑2/E
1	1	3	2.86	.0071
1	2	7	4.76	1.0519
1	3	8	5.95	.7044
1	4	2	6.43	3.0508
2	1	4	3.43	.0952
2	2	5	5.71	.0893
2	3	9	7.14	.4829
2	4	6	7.71	.3810
3	1	3	2.43	.1345
3	2	1	4.05	2.2947
3	3	2	5.06	1.8501
3	4	11	5.46	5.6081
4	1	2	3.29	.5031
4	2	7	5.48	.4240
4	3	6	6.85	.1044
4	4	8	7.39	.0499

Note that the third column gives the observed frequencies n_{ij}, which may be used to check data input if this is not done in (B). The expected values are computed as

$$E_{ij} = R_i C_j / N$$

and the last column gives the "contributions to chi-square",

$$(O_{ij} - E_{ij})^2 / E_{ij} \quad .$$

4. CHI-SQUARE

For the example table this part of the output appears as follows:

```
CHI-SQUARE =  16.831 WITH 9 DEGREES OF FREEDOM
(SMALLEST EXPECTED VALUE:  2.43)
```

Chi-square is obtained by adding the "contributions" displayed in the last column of the output from (3), and the degrees of freedom are (I-1)(J-1). The smallest expected value is always given, as a help in deciding how much faith to put in the accuracy of the chi-square approximation.

If any column of the table is empty the output appears thus instead:

```
CHI-SQUARE NOT COMPUTED
(SMALLEST EXPECTED VALUE:  0.00)
```

The correct chi-square for the non-empty columns can be obtained by adding the "contributions" given in the last column of the output from (3), and the degrees of freedom will be (I-1)(J'-1) where J' is the number of non-empty columns. Or, if convenient, the table may just be re-entered with the empty columns omitted.

5. FISHER'S TEST

This is computed for 2x2 tables only (thus not for the example), if $\min(R_1, R_2, C_1, C_2) \leq 50$. (When all row and column totals exceed 50 the chi-square approximation should be adequate.)

Here is given the P-value, or level of significance, for testing independence against the one-sided alternative of positive association, using the Fisher-Irwin "exact" test procedure. This is appropriate if both margins of the table were fixed in advance, or if a conditional test, given the margins, is desired. The formula is

$$\text{P-positive} = \sum_{k=n_{11}}^{\min(R_1,C_1)} \frac{R_1!\ R_2!\ C_1!\ C_2!}{N!\ k!\ (R_1-k)!\ (C_1-k)!\ (N-R_1-C_1+k)!} .$$

To obtain a P-value (say "P-negative") for testing against negative association, re-enter the table with rows or columns (not both!) reversed.

For a two-sided P-value, against positive or negative association, use the upper bound P-absolute $\leq 2 \times \min$(P-positive, P-negative). This upper bound is exact if $R_1=R_2$ or if $C_1=C_2$ or both. Note that P-positive and P-negative cannot both be less than ½; hence, if P-positive is $\leq$ ½ it is the minimum of the two and twice its value is an upper bound on P-absolute, the computation of P-negative then being unnecessary.

6. (OPTION) TESTS OF SYMMETRY

(This option is available only for square tables: i.e. if I = J = M, say.) For the example table the output for this option, if it is taken, will appear as follows:

```
DO YOU WANT TO TEST SYMMETRY?
?YES

FOR TESTING SYMMETRY AGAINST GENERAL ALTERNATIVES:
BOWKER'S CHI-SQUARE = 22.077 WITH 6 DEGREES OF FREEDOM
(SMALLEST EXPECTED VALUE:  2.00)
```

```
FOR TESTING SYMMETRY AGAINST 'DIAGONAL SKEWNESS':
SEN'S CHI-SQUARE = 14.807 WITH 3 DEGREES OF FREEDOM
(SMALLEST EXPECTED VALUE:  2.00)

FOR SIGN TEST OF STOCHASTIC EQUALITY OF MARGINALS:
APPROXIMATE NORMAL DEVIATE:  2.339
(43 OBSERVATIONS ABOVE DIAGONAL AND 23 BELOW)
```

The hypothesis of symmetry for a bivariate categorical random variable with M categories is that $p_{ij} = p_{ji}$ for $1 \leq i < j \leq M$, where p_{ij} is the probability for the (i,j) cell.

6.1 Bowker's chi-square (omitted if I = J = 2)

The formula is:

$$X_B^2 = 2 \sum_{i<j} \sum \frac{(n_{ij} - n_{ji})^2}{n_{ij} + n_{ji}} ,$$

with M(M-1)/2 degrees of freedom. The minimum expected value is

$$\min_{i<j} \frac{n_{ij} + n_{ji}}{2}$$

and if this equals zero then X_B^2 is not computed. Bowker's chi-square may be used to test symmetry against the general alternative that $p_{ij} \neq p_{ji}$ for some $i < j$, $1 \leq i < j \leq M$.

6.2 Sen's chi-square (omitted if I = J = 2)

The formula is:

$$X_S^2 = 2 \sum_{k=1}^{M-1} \frac{(F_k - F_{-k})^2}{F_k + F_{-k}} ,$$

with (M-1) degrees of freedom, where

$$F_k = \begin{cases} \sum_{u=1}^{M-k} n_{u+k,u} & \text{for } -(M-1) \leq k \leq -1 \\ \sum_{u=1}^{M-k} n_{u,u+k} & \text{for } 1 \leq k \leq (M-1). \end{cases}$$

The minimum expected value is

$$\min_{1 \leq k \leq M-1} \frac{F_k + F_{-k}}{2}$$

and if this equals zero then X_S^2 is not computed. Sen's chi-square may be used to test symmetry against the alternative ("diagonal skewness") that $P_k \neq P_{-k}$ for some k, where

$$P_k = \begin{cases} \sum_{u=1}^{M-k} p_{u+k,u} & \text{for } -(M-1) \leq k \leq -1 \\ \sum_{u=1}^{M-k} p_{u,u+k} & \text{for } 1 \leq k \leq (M-1). \end{cases}$$

For further discussion see P. K. Sen, "On some nonparametric tests for symmetry in twoway tables", Journal of the Indian Statistical Association, Vol. 4 (1966).

6.3 Sign test

This is just a comparison of the number of observations above the diagonal (say, A) with the number of observations below the diagonal (say, B); the approximate normal deviate is

$$Z = \begin{cases} 0 & \text{if } A = B \\ \frac{|A-B|-1}{\sqrt{A+B}} & \text{if } A \neq B. \end{cases}$$

For a 2x2 table this test is sometimes known as the McNemar test for the significance of changes. It may be thought of as testing the equality of the marginals of the bivariate distribution against an alternative of difference in location.

7. (OPTION) COMPARISON OF ROWS

This option is not available if there are only two columns (J=2), and it is not meaningful unless the ordering of the columns is relevant. Note that a comparison of columns is not available as such, but can be obtained by re-entering the table with rows and columns transposed.

For the example table, the output for this option would appear as follows:

```
DO YOU WANT TO COMPARE ROWS, ASSUMING COLUMN ORDERING IS RELEVANT?
?YES

ROW          N      MEAN     MEDIAN    PROBABILITY EFFECT

 1          20      2.450     2.500      -.204
 2          24      2.708     2.833      -.050
 3          17      3.235     3.727       .267
 4          23      2.870     2.917       .032

TOTAL       84      2.798     2.900


MEDIAN TEST
     CHI-SQUARE =  2.745 WITH 3 DEGREES OF FREEDOM

RANK ANALYSIS OF VARIANCE (KRUSKAL-WALLIS TEST)
     F =  2.384 WITH 2.96 AND 78.85 DEGREES OF FREEDOM

STANDARD ANALYSIS OF VARIANCE (ASSUMING EQUALLY-SPACED COLUMNS)
     F =  1.865 WITH 3 AND 80 DEGREES OF FREEDOM
```

7.1 Means

For $1 \leq i \leq I$, the mean of the i-th row is

$$\bar{x}_i = \sum_{j=1}^{J} j \, n_{ij}/R_i \quad ;$$

the overall mean is

$$\bar{\bar{x}} = \sum_{j=1}^{J} j \, C_j/N \; .$$

These means have no reasonable interpretation unless it is assumed that the columns are equally spaced along some scale.

7.2 Medians

The medians are calculated on the assumption that the columns resulted from grouping continuous data, and that in particular any column — say, the j-th — represents the interval from $(j-\frac{1}{2})$ to $(j+\frac{1}{2})$. (If the columns actually represent a discrete variable then each median as given may be rounded to the nearest integer.)

Warning: if the median of any row lies inside an empty interval, or between two empty intervals, it will be printed out incorrectly as the upper end of the last preceding nonempty interval. For example, in the table

1	2	2	0	5
4	0	0	2	2

the correct row medians are 4.0 and 2.5 but they would appear as 3.5 and 1.5. Similar remarks hold for the overall median. (This does not affect the median test.)

7.3 Probability effects

Let X_i be an observation chosen at random from the population represented by the i-th row, say, and let X be an observation chosen at random from the overall population (or the mixture of row populations in which the proportions are as in the table at hand). Then the true probability effect of the i-th row is

$$\theta_i = P\{X_i > X\} - P\{X_i < X\} .$$

(These are genuine "effects",in that $\Sigma R_i \theta_i = 0$.)

Let Y_j be the column-rank of any observation in the j-th column, computed as

$$Y_j = \sum_{j'<j} C_{j'} + \tfrac{1}{2}(C_j+1) \quad \text{for } 1 \leq j \leq J \text{ ;}$$

and let

$$\bar{X}_i = \sum_{j=1}^{J} Y_j \, n_{ij}/R_i$$

be the mean rank in the i-th row. Then an unbiased estimate of θ_i is

$$\hat{\theta}_i = 2\bar{X}_i - 1 \quad ,$$

and this is what the computer prints out.

7.4 <u>Median test</u>

This is the chi-square test for testing homogeneity of rows in the table which results after the J columns have been collapsed into two in such a way as to make the two column totals as nearly equal as possible (without re-ordering). For the example, the resulting table being tested is as follows:

10	10	20
9	15	24
4	13	17
9	14	23
32	52	84

7.5 <u>Rank analysis of variance</u>

Rank analysis of variance is the appropriate procedure for testing homogeneity of rows against the alternative that at least one probability effect is non-zero. (If the hypothesis is true, then the probability effects will all be zero, although the converse does not necessarily hold.) The test statistic is

$$F = \frac{N-I}{I-1} \; \frac{\sum_{i=1}^{I} R_i \bar{X}_i^2 - n(n+1)^2/4}{\sum_{i=1}^{I} \sum_{j=1}^{J} Y_j^2 n_{ij} - \sum_{i=1}^{I} R_i \bar{X}_i^2} ,$$

where $\bar{X}_i$ and Y_j are as defined in (7.3) above. This is related to the more commonly used Kruskal-Wallis H-statistic (when the latter is corrected for ties in the customary manner) by the formula

$$H = \frac{(N-1)(I-1)F}{(N-I) + (I-1)F} .$$

The usual Kruskal-Wallis test is based on treating H as a chi-square variable with (I-1) degrees of freedom. However, the program uses the improved approximation obtained by treating F as an F-variable with (I-1)D and (N-I)D degrees of freedom, where

$$D = 1 - \frac{6(N+1)}{(N-1)(5N+6)} .$$

This is the "B_2 approximation" suggested by David L. Wallace in the Journal of the American Statistical Association, Vol. 54 (1959); it should be quite satisfactory if all row totals are at least 5 or more.

7.6 Standard analysis of variance

Standard analysis of variance is the appropriate procedure for testing homogeneity of rows against the alternative that the row means are not all equal. (If the hypothesis is true then the row means will all be equal, although the converse does not necessarily hold.) The test statistic is

$$F = \frac{N-I}{I-1} \; \frac{\sum_{i=1}^{I} R_i \bar{x}_i^2 - N \bar{\bar{x}}^2}{\sum_{i=1}^{I} \sum_{j=1}^{J} j^2 n_{ij} - \sum_{i=1}^{I} R_i \bar{x}_i^2} ,$$

where $\bar{\bar{x}}$ and the $\bar{x}_i$'s are the means as defined in (7.1) above. The statistic F has asymptotically, under the hypothesis of homogeneity, an F-distribution with (I-1) and (N-I) degrees of freedom. (If column subscripts were exactly normally distributed — which is obviously not possible — then the asymptotic distribution would be exact even for small N.) The test has no reasonable interpretation unless it is assumed that the columns are equally spaced along some scale.

8. (OPTION) CORRELATION ANALYSIS

This option is, of course, not meaningful unless the orderings of the rows and columns are both relevant. For the example table the output would appear as follows:

```
DO YOU WANT A CORRELATION ANALYSIS?
?YES

TOTAL PAIRS:       3486.
CONCORDANT:        1198.
DISCORDANT:         770.
ROW TIES:           855.
COLUMN TIES:        907.
```

KENDALL CORRELATION (TAU-B): .1643
NORMAL DEVIATE FOR TEST OF INDEPENDENCE: 1.7834

UNCONDITIONAL INDEX (TAU-A): .1228
STANDARD ERROR: .0616

GOODMAN-KRUSKAL (CONDITIONAL) INDEX (G): .2175
STANDARD ERROR: .1087

SPEARMAN RANK CORRELATION: .1944

PRODUCT-MOMENT CORRELATION: .1789
(ASSUMING EQUALLY-SPACED ROWS AND COLUMNS)

8.1 Counts of pairs

The total number of pairs is

$$NP = \tfrac{1}{2} N(N-1) .$$

Let $C_{ij} = \sum_{i'<j} \sum_{j'<j} n_{i'j'} + \sum_{i'>i} \sum_{j'>j} n_{i'j'}$

and $D_{ij} = \sum_{i'<i} \sum_{j'>j} n_{i'j'} + \sum_{i'>i} \sum_{j'<j} n_{i'j'}$

for $1 \leq i \leq I$, $1 \leq j \leq J$. Then C_{ij} is the number of observations concordant with any one observation in the (i,j) cell, and the number of concordant pairs is

$$NC = \tfrac{1}{2} \sum_{ij}\sum n_{ij} C_{ij} .$$

Similarly, D_{ij} is the number of observations discordant with any one observation in the (i,j) cell, and the number of discordant pairs is

$$ND = \tfrac{1}{2} \sum_{ij}\sum n_{ij} D_{ij} .$$

The number of row ties is

$$NX = \tfrac{1}{2} \sum_{i} R_i (R_i - 1)$$

and the number of column ties is

$$NY = \tfrac{1}{2} \sum_{j} C_j (C_j - 1) .$$

The total number of ties, eliminating the duplicate "double ties", may be calculated as

$$NT = NP - NC - ND;$$

this is not printed out, however.

8.2 Kendall correlation

The Kendall Correlation coefficient, or "tau-b", is calculated as

$$\text{tau-b} = \frac{NC - ND}{\sqrt{(NP-NX)(NP-NY)}} \quad .$$

A test of independence based on this coefficient uses

$$Z = 3(NC-ND)\sqrt{\frac{N(N-1)(N-2)}{Q_1Q_2 + 18(N-2)(NP-NX)(NP-NY)}}$$

as a normal deviate, where

$$Q_1 = N(N-1)(N-2) - \sum_i R_i(R_i-1)(R_i-2)$$

and

$$Q_2 = N(N-1)(N-2) - \sum_j C_j(C_j-1)(C_j-2) \quad .$$

This value of Z is fully adjusted for ties. The approximation to normality may be regarded as reasonably accurate if $N >= 20$, excellent if $N >= 50$.

8.3 Unconditional index

The unconditional index of order-association, or Kendall's "tau-a", is

$$\text{tau-a} = \frac{NC-ND}{NP} \quad .$$

Its standard error is

$$S_a = \frac{1}{NP}\sqrt{\Sigma\Sigma n_{ij}(C_{ij}-D_{ij})^2 - 4(NC-ND)^2/N} \quad .$$

8.4 Goodman-Kruskal or conditional index

The conditional index of order-association, or Goodman and Kruskal's "G", is

$$G = \frac{NC-ND}{NC+ND}$$

and its standard error is

$$S_G = \left(\frac{2}{NC+ND}\right)^2 \sqrt{(ND)^2\Sigma\Sigma n_{ij}C_{ij}^2 - 2(NC)(ND)\Sigma\Sigma n_{ij}C_{ij}D_{ij} + (NC)^2\Sigma\Sigma n_{ij}D_{ij}^2} \ .$$

8.5 Spearman rank correlation

(This is not printed out if $I = J = 2$, in which case it would be identical with the Kendall correlation.)

Define

$$X_i = \sum_{i'<i} R_{i'} + \tfrac{1}{2}(R_i+1) \qquad \text{for } 1 \le i \le I$$

and

$$Y_j = \sum_{j'<j} C_{j'} + \tfrac{1}{2}(C_j+1) \qquad \text{for } 1 \le j \le J;$$

X_i may be interpreted as the row-rank of any observation in the i-th row and Y_j as the column-rank of any observation in the j-th column. Then the sample Spearman rank correlation between rows and columns is

$$r_S = \frac{\Sigma\Sigma(X_i-\bar{X})(Y_j-\bar{Y})n_{ij}}{\sqrt{\Sigma(X_i-\bar{X})^2R_i}\sqrt{\Sigma(Y_j-\bar{Y})^2C_j}} \ ,$$

where

$$\bar{X} = \sum_{i=1}^{I} X_iR_i/N \qquad \text{and} \qquad \bar{Y} = \sum_{j=1}^{J} Y_jC_j/N \ .$$

(This is just the ordinary product-moment correlation between row-ranks and column-ranks.)

8.6 Product-moment correlation

(This is not printed out if I = J = 2, in which case it would be identical with the Kendall Correlation.) The sample product-moment correlation between rows and columns is

$$r = \frac{\Sigma\Sigma(i-\bar{i})(j-\bar{j})n_{ij}}{\sqrt{\Sigma(i-\bar{i})^2 R_i}\ \sqrt{\Sigma(j-\bar{j})^2 C_j}} ,$$

where

$$\bar{i} = \sum_{i=1}^{I} iR_i/N \quad \text{and} \quad \bar{j} = \sum_{j=1}^{J} jC_j/N \quad .$$

This measure of correlation has no reasonable interpretation unless it is assumed that the rows and the columns both represent equally-spaced values on corresponding scales.

9. OPTION TO REPEAT SAME ANALYSIS WITH DIFFERENT DATA

When the analysis of the contingency table is completed, the computer asks:

FOR YOUR NEXT TABLE, DO YOU WANT TO CHANGE DIMENSIONS OR OPTIONS?

If the user answers "YES", the computer returns to (1.) and the program begins again as if it had just been called.

If the user answers "NO", the computer understands this to mean that the next table has exactly as many rows and columns as the one just analyzed, and that the same options are desired. It then goes directly to the request

ENTER ROW 1

and demands the same amount of data as before, and performs the same analysis without asking further questions. This option makes the work go much faster when the user has a series of similar tables to analyze.

If the user has no more tables to analyze, he should answer "STOP".

There follows another example of the output from the program. This time a 2x2 table has been used, in order to illustrate the printout for Fisher's test and the curtailment of the other analyses.

```
THIS PROGRAM ANALYZES TWO-WAY CONTINGENCY TABLES.

HOW MANY ROWS (2 TO 50) AND COLUMNS (2 TO 25) ?
?2 2

ENTER ROW 1
?39 16

ENTER ROW 2
?21 34

DO YOU WANT THE TABLE PRINTED OUT?
?YES

  R\C       1       2     TOTAL

   1       39      16        55

   2       21      34        55

TOTAL      60      50       110


DO YOU WANT EXPECTED VALUES?
?YES


ROW  COLUMN  OBSERVED  EXPECTED  (O-E)↑2/E

  1      1         39     30.00     2.7000
  1      2         16     25.00     3.2400
  2      1         21     30.00     2.7000
  2      2         34     25.00     3.2400


CHI-SQUARE =  11.880 WITH 1 DEGREES OF FREEDOM
(SMALLEST EXPECTED VALUE:  25.00)

FISHER P-VALUE AGAINST POSITIVE ASSOCIATION:  .000518


DO YOU WANT TO TEST SYMMETRY?
?YES

FOR SIGN TEST OF STOCHASTIC EQUALITY OF MARGINALS:
APPROXIMATE NORMAL DEVIATE:  .658
(16 OBSERVATIONS ABOVE DIAGONAL AND 21 BELOW)
```

```
DO YOU WANT A CORRELATION ANALYSIS?
?YES


TOTAL PAIRS:      5995.
CONCORDANT:       1326.
DISCORDANT:        336.
ROW TIES:         2970.
COLUMN TIES:      2995.

KENDALL CORRELATION (TAU-B):  .3286
NORMAL DEVIATE FOR TEST OF INDEPENDENCE:  3.4310

UNCONDITIONAL INDEX (TAU-A):  .1651
STANDARD ERROR:  .0453

GOODMAN-KRUSKAL (CONDITIONAL) INDEX (G):  .5957
STANDARD ERROR:  .1311


FOR YOUR NEXT TABLE, DO YOU WANT TO CHANGE DIMENSIONS OR OPTIONS?
?STOP

RAN 42 SEC.
```

APPENDIX: LISTING OF THE PROGRAM

```
10 DIMENSION N(50,25), NR(50), NC(25); 50 FORMAT(/"ENTER ROW ",I1)
20 PRINT"THIS PROGRAM ANALYZES TWO-WAY CONTINGENCY TABLES."
30 80S=1;81PRINT,↑"HOW MANY ROWS (2 TO 50) AND COLUMNS (2 TO 25) ?"
40 INPUT,LI,LJ;IF(LI-2)81;IF(50-LI)81;IF(LJ-2)81;IF(25-LJ)81;82U=1;G=0
50 DO24J=1,LJ;24NC(J)=0;DO6I=1,LI;NR(I)=0;38PRINT50,I;DO5J=1,LJ;INPUT,M
60 IF(M)31;G=G+M;IF(524287-G)32;NR(I)=NR(I)+M;NC(J)=NC(J)+M;5N(I,J)=M
70 IF(NR(I)-1)33;6;IF(S)25;PRINT,↑"DO YOU WANT THE TABLE PRINTED OUT?"
80 INPUT,K1;25IF(K1-"NO")19,27,19;27PRINT,↑"N =",;GOTO36
90 19PRINT,↑"  R\C",;PRINT11,(J,J=1,LJ),;PRINT"    TOTAL"
100 DO12I=1,LI;PRINT10,(I,(N(I,J),J=1,LJ)),;PRINT"  ",;12PRINT11,NR(I)
110 PRINT,↑"TOTAL",;PRINT11,(NC(J),J=1,LJ),;PRINT"  ",;36PRINT11,G;IF(S)9
120 28PRINT,↑↑"DO YOU WANT EXPECTED VALUES?";INPUT,K2;9IF(K2-"NO")84,1
130 84PRINT,↑↑"ROW  COLUMN  OBSERVED  EXPECTED  (O-E)↑2/E",↑;GOTO85;1U=-1
140 85Q=G;X=0;DO91I=1,LI;DO91J=1,LJ;E=NR(I)*(NC(J)/G);IF(U)87
150 PRINT86,I,J,N(I,J),E,;86FORMAT(I3,I8,I10,F10.2);87IF(E)90,90
160 W=((N(I,J)-E)↑2)/E;X=X+W;90IF(U)91;IF(E)89,89;PRINT88,W,
170 89PRINT;91Q=MIN1F(E,Q);DF=(LI-1)*(LJ-1);PRINT,↑↑"CHI-SQUARE",
180 IF(-E)92;PRINT78;GOTO93;92PRINT77,X,DF;93PRINT79,Q
190 79FORMAT("(SMALLEST EXPECTED VALUE: ",F3.2,")",/)
200 IF(1-DF)60;A=MINOF(NC,NR);B=MAXOF(NC,NR);C=MINOF(NC(2),NR(2));L=A-N
210 IF(A-C)43;A=C;B=MAXOF(NC(2),NR(2));L=A-N(2,2);43IF(50-A)4;H=Q=T=1
220 DO44K=1,A;44H=H*(B-A+K)/(G-A+J);IF(L-1)46;DO45K=1,L
230 Q=Q*(((A-K+1)*(G-B-K+1))/(K*(B-A+K)));45T=T+Q;46P=H*T;PRINT52,P
240 52FORMAT(/"FISHER P-VALUE AGAINST POSITIVE ASSOCIATION: ",F7.6,/)
250 60IF(LI-LJ)3,4,3;4IF(S)61;PRINT,↑"DO YOU WANT TO TEST SYMMETRY?"
260 INPUT,K3;61IF(K3-"NO")62,3;62IF(LI-3)71;PRINT,↑"FOR TESTING SYMMETRY
270 + AGAINST GENERAL ALTERNATIVES:";PRINT"BOWKER'S CHI-SQUARE",;X=0;E=G
280 DO64I=2,LI;K=I-1;DO64J=1,K;Q=N(I,J)+N(J,I);E=MIN1F(E,Q);IF(Q)65,65
290 64X=X+((N(I,J)-N(J,I))↑2)/Q;PRINT77,2*X,LI*(LI-1)/2;GOTO66;65PRINT78
300 66PRINT79,E/2.;78FORMAT(" NOT COMPUTED");7FORMAT(F5.3)
310 PRINT,↑"FOR TESTING SYMMETRY AGAINST 'DIAGONAL SKEWNESS':"
320 PRINT"SEN'S CHI-SQUARE",;X=0;E=G;DO68M=2,LI;A=B=0;K=LI-M+1;DO67J=1,K
330 A=A+N(J,J+M-1);67B=B+N(J+M-1,J);Q=A+B;E=MIN1F(E,Q);IF(Q)69,69
340 68X=X+((A-B)↑2)/Q;PRINT77,2*X,LI-1;GOTO70;69PRINT78;70PRINT79,E/2
350 71PRINT,↑"FOR SIGN TEST OF STOCHASTIC EQUALITY OF MARGINALS:"
360 PRINT"APPROXIMATE NORMAL DEVIATE: ",;A=B=W=0;DO73I=1,LI;DO73J=1,LJ
370 Q=N(I,J);IF(J-I)72,73;A=A+Q;GOTO73;72B=B+Q;73;Q=SQRT(A+B);IF(-Q)74
380 PRINT78;GOTO76;74Z=ABS(A-B)-1;IF(Z)75;W=Z/Q;75PRINT7,W;76PRINT8,A,B
390 8FORMAT("(",I1," OBSERVATIONS ABOVE DIAGONAL AND ",I1," BELOW)")
400 77FORMAT(" = ",F4.3," WITH ",I1," DEGREES OF FREEDOM")
410 3IF(LJ-3)39;IF(S)40;PRINT,↑↑"DO YOU WANT TO COMPARE ROWS, ASSUMING
420 + COLUMN ORDERING IS RELEVANT?";INPUT,K4;40IF(K4-"NO")49,39
430 49PRINT,↑"ROW        N     MEAN    MEDIAN    PROBABILITY EFFECT",↑
440 Q=1;R=(G-LI)/(LI-1);A=E=F=T=U=V=W=0;DO51J=1,LJ;H=NC(J);W=W+H
450 IF(Q)51;IF(W-G/2)51;Q=-1;C=J+.5-(W-G/2)/H;B=W;IF(J-C)51
460 B=B-H;51A=A+J*H;DO59I=1,LI;Q=1;P=W=X=Y=Z=0;H=NR(I);DO58J=1,LJ
470 O=N(I,J);W=W+.5*NC(J);Y=Y+O*W;F=F+O*W*W;W=W+.5*NC(J);E=E+O*J*J
480 X=X+O*J;P=P+O;IF(C-J)57;Z=Z+O;57IF(Q)58;IF(P-H/2)58;Q=-1
490 D=J+.5-(P-H/2)/O;58;T=T+X*X/H;U=U+Y*Y/H;V=V+((G*Z-B*H)↑2)/H
```

```
59PRINT63,I,H,X/H,D,(2*Y)/(G*H)-1;D=1-(6*G+6)/(5*G*G+G-6)
PRINT99,G,A/G,C,V/(B*(G-B)),LI-1,R*(U-.25*G↑3)/(F-U),
+D*(LI-1),D*(G-LI),R*(T-A*A/G)/(E-T),LI-1,G-LI
99FORMAT(/"TOTAL",I9,2F10.3,///"MEDIAN TEST",/5X"CHI-SQUARE = ",
+F4.3," WITH ",I1," DEGREES OF FREEDOM",//"RANK ANALYSIS OF VARIANCE
+ (KRUSKAL-WALLIS TEST)",/5X"F = ",F4.3," WITH ",F3.2," AND ",F3.2,
+" DEGREES OF FREEDOM",//"STANDARD ANALYSIS OF VARIANCE
+ (ASSUMING EQUALLY-SPACED COLUMNS)",/5X"F = ",F4.3," WITH ",
+I1," AND ",I1," DEGREES OF FREEDOM");88FORMAT(F11.4)
39IF(S)47;PRINT,↑↑"DO YOU WANT A CORRELATION ANALYSIS?"
INPUT,K5;47IF(K5-"NO")34,48;34SC=SD=SCC=SCD=SDD=0;P=G*(G-1)/2
DO15I=1,LI;DO15J=1,LJ;C=D=0;10FORMAT(/I4,X,8I7,3(/5X,8I7))
DO14IP=1,LI;DO14JP=1,LJ;IF((I-IP)*(J-JP))13,14;C=C+N(IP,JP);GOTO14
13D=D+N(IP,JP);14;20FORMAT("STANDARD ERROR: ",F5.4)
SCC=SCC+N(I,J)*C*C;SCD=SCD+N(I,J)*C*D;SDD=SDD+N(I,J)*D*D
SC=SC+N(I,J)*C;15SD=SD+N(I,J)*D;PC=SC/2;PD=SD/2;QQ=2*P*(G-2)
TR=0;QR=QQ;DO16I=1,LI;W=NR(I);TR=TR+W*(W-1)/2;16QR=QR-W*(W-1)*(W-2)
TC=0;QC=QQ;DO17J=1,LJ;W=NC(J);TC=TC+W*(W-1)/2;17QC=QC-W*(W-1)*(W-2)
PRINT,↑↑"TOTAL PAIRS: ",;PRINT29,P;29FORMAT(F9.0)
PRINT"CONCORDANT:  ",;PRINT29,PC;PRINT"DISCORDANT:  ",;PRINT29,PD
PRINT"ROW TIES:     ",;PRINT29,TR;PRINT"COLUMN TIES: ",;PRINT29,TC
IF(PC+PD)48,48;TB=(PC-PD)/SQRT((P-TR)*(P-TC));18FORMAT(F5.4)
Z=3.*(PC-PD)*SQRT(QQ/(QR*QC+18.*(G-2)*(P-TR)*(P-TC)))
SA=SQRT(SCC-2*SCD+SDD-((SC-SD)↑2)/G)/P;11FORMAT(8I7,3(/5X,8I7))
SG=((2/(SC+SD))**2)*SQRT(SCC*SD*SD-2*SC*SD*SCD+SDD*SC*SC)
PRINT,↑"KENDALL CORRELATION (TAU-B): ",;PRINT18,TB
PRINT"NORMAL DEVIATE FOR TEST OF INDEPENDENCE: ",;PRINT18,Z
PRINT,↑"UNCONDITIONAL INDEX (TAU-A): ",;PRINT18,(PC-PD)/P
PRINT20,SA;PRINT,↑"GOODMAN-KRUSKAL (CONDITIONAL) INDEX (G): ",
PRINT18,(PC-PD)/(PC+PD);PRINT20,SG;IF(DF-2)48
A=B=C=0;X=-G;DO23I=1,LI;X=X+NR(I);Y=-G;DO22J=1,LJ;Y=Y+NC(J)
Q=N(I,J);A=A+Q*X*X;B=B+Q*Y*Y;C=C+Q*X*Y;22Y=Y+NC(J);23X=X+NR(I)
PRINT,↑"SPEARMAN RANK CORRELATION: ",;PRINT18,C/SQRT(A*B)
X=Y=A=B=C=0;DO53U=1,LI;53X=X+U*NR(U);X=X/G;DO54V=1,LJ;54Y=Y+V*NC(V)
Y=Y/G;DO55I=1,LI;DO55J=1,LJ;Z=N(I,J);A=A+Z*(I-X)↑2;B=B+Z*(J-Y)↑2
55C=C+Z*(I-X)*(J-Y);PRINT56,C/SQRT(A*B);56FORMAT(/"PRODUCT-MOMENT
+ CORRELATION: ",F5.4,/"(ASSUMING EQUALLY-SPACED ROWS AND COLUMNS)")
48PRINT,↑↑"FOR YOUR NEXT TABLE, DO YOU WANT TO CHANGE DIMENSIONS OR
+ OPTIONS?";INPUT,K;IF(K-"NO")80,37,80;63FORMAT(I2,I12,3F10.3)
37S=-1;GOTO82;31STOP"NEG";32STOP"BIG";33STOP"ROW";END
```