United States
Department of
Agriculture

**Forest
Service**

**North Central
Research Station**

**General Technical
Report NC-232**

TableSim—A Program for Analysis of Small-Sample Categorical Data

David J. Rugg

Disclaimer

The computer program described in this publication is available on request with the understanding that the U.S. Department of Agriculture cannot assure its accuracy, completeness, reliability, or suitability for any other purpose than that reported. The recipient may not assert any proprietary rights thereto nor represent it to anyone as other than a Government-produced computer program.

Table of Contents

Download self-extracting zip file containing manual, software, source code, and examples from :
http://www.ncrs.fs.fed.us/pubs/viewpub.asp?key=1838

Published by:
North Central Research Station
Forest Service U.S. Department of Agriculture
1992 Folwell Avenue
St. Paul, MN 55108
2003

Web site: www.ncrs.fs.fed.us

TableSim—A Program for Analysis of Small-Sample Categorical Data

A rich set of analysis tools exists for categorical data when sample sizes are large. However, due to constraints in the frequency of the phenomenon under study or resources available for gathering data, researchers often find themselves with sample sizes too small for reliable inference using asymptotic methods. And, unfortunately, they have no good guides to tell them when a data set is "too small" for asymptotic methods—sample size interacts with table structure in unpredictable ways. Expected values less than five commonly trigger sample size warnings in standard statistical software (e.g., Systat). However, there is also evidence that expected cell sizes can be as small as 1.0 without adversely affecting the P-values (Fienberg 1980, Appendix IV). Historically, very few alternatives to asymptotic analyses were available, so researchers have had to hope for the best. However, as microcomputers became faster, the use of randomization methods became feasible. Research in efficient computation of exact distributions brought further improvements. For example, the commercial software program StatXact (Mehta and Patel 1995) performs many small sample-analyses either exactly or using randomization methods.

TableSim, the program described in this manual, performs four types of small-sample categorical data analysis.

1. Test of homogeneity or independence for r x c tables.
2. Test of homogeneity of an r x c table across k strata.
3. Comparison of an observed r x c table to a specified distribution.
4. Test of homogeneity or independence for r x c tables with repeated measures (aggregated data) in the rows. This module is also useful for large-sample inference.

Of these analyses, only the first is directly available in StatXact.

In this manual I will describe how to make TableSim correctly execute its functions and how the program does its work. Examples are provided for each function, showing how to set up and run the analyses and how to interpret the output. Finally, some guidance is provided for making the source code run under operating systems other than Windows.

The primary audience for this program is statisticians, researchers, and technicians who are already familiar with standard asymptotic categorical data methods, but who need to apply those methods to small-sample data sets. This publication is not a primer on categorical data analysis. Excellent introductions to categorical data analysis can be found in Fienberg (1980) and Agresti (1996). The secondary audience is programmers making modifications to the original program.

About The Author:

David J. Rugg is Mathematical Statistician with the USDA Forest Service, North Central Research Station, 1992 Folwell Ave., St. Paul, MN 55108; Phone: (651) 649-5173; e-mail: drugg@fs.fed.us.

Program Documentation

TableSim runs from the command prompt of any 32-bit Microsoft® Windows® operating system (see User Note 1). Being a command line program, TableSim requires no special installation—simply copy the file “tablesim.exe” into a directory. This installation method, while simple and flexible, provides no program icon in the “Start” menu. To minimize the amount of path information to be typed in, a copy of the program can be put in the directory holding the data being worked on.

The size of problem handled by these routines is limited primarily by the amount of RAM available on the computer. The secondary limit is that the program will only read or write lines that are shorter than 500 characters, which affects the number of columns in a table. (Modifying this is discussed in the Source Code Issues section.) No constraints on the number of rows, columns, or tables are hard-coded into the software. All testing of the program was performed under Windows 2000 Professional.

File Structures

General input file structure

While most of the analyses have some special requirements for their input files, as described in the analysis sections below, the input file generally has this structure:

- Line 1 is the analysis label; it must be ≤ 100 characters and enclosed in quotes.
- Line 2 is the analysis keyword, and must be enclosed in quotes.
- Line 3 is the dimension of the problem.
- Line 4 is the number of random tables the program uses to estimate the P-value.
- Line 5 is blank.
- The next lines contain the observations; these must be integers.
- There is a blank line after the last row of observations.

There are some general rules for formatting input files.

- All values on a particular line of an input file must be separated by at least one space, as shown in the figures referenced in the sections below.
- Quotation marks can be single or double as long as they are used consistently on a particular line.
- Blank lines require only a carriage return.
- The number of simulations to run must always be specified. The program will execute properly using any integer from 1 to 2,147,483,647.

A common question is how many simulations to run; 10,000 is a reasonable minimum. More will be needed if the P-value needs to be determined to a very high precision (recall that the variance of the P-value is P*(1-P)/N, where N = number of simulations). More simulations are also needed when the P-value is very small, and the analyst wants to know just how small, rather than simply reporting, say, $P < 0.0001$. The smallest P-value the program can report is 1/N, where N is the value specified in the input file. The observed data form the first simulation, consistent with standard practice for randomization tests (Manly 1997, page 7).

General output file structure

The general output file structure is:

- the user-provided analysis label;
- observed data structure as a check that the program worked on the desired data, and as an additional reminder about what the analysis pertains to;
- expected data structure under the null hypothesis;
- standardized residual matrices;
- Pearson X^2 and likelihood ratio G^2 statistics, and their Monte Carlo P-values.

The standardized residual matrices (Agresti 1990, page 224) are computed using the formula:

(observed count – expected count)/(square root of expected count).

Like regression residuals, standardized residuals show where the model that generated the expected values is failing to explain the observed values. Clearly, positive residuals mean that more observations were recorded than expected, and negative residuals mean that fewer observations were recorded than expected. The formula should be familiar to most users—sum the squared residuals, and the result is the Pearson X^2 statistic.

Data Analysis Routines

Figure 1a and figure 1b show two example data analysis runs of TableSim. The program is invoked, and it asks for the name of an input file. After ensuring that the input file exists, the program requests the name of an output file. The specified output file is created or the user is asked whether to overwrite the existing file with that name. If the user chooses not to overwrite, the program asks for a new output file name. Given the file information, the program executes and terminates.

Independence in r x c tables

This is the standard analysis for two-dimensional, unordered data. Figure 2a shows a sample input file; figure 2b shows the resulting output file.

Homogeneity of an r x c table across k strata

This analysis is a special extension of independence into three dimensions. The cell probabilities in the r x c table are of general form—neither independence nor any other structure is assumed. In principle, the analysis constructs a pooled r x c table and then compares the k observed tables to their expected values based on the pooled table. StatXact's test for strata is not comparable to this analysis. However, StatXact will run this analysis indirectly—set up a 2-way table with k rows and r x c columns and request a test of independence. TableSim employs this structure internally, but allows the user to set up the problem with the more obvious structure. Figure 3a shows a sample input file for this analysis; figure 3b shows the resulting output file.

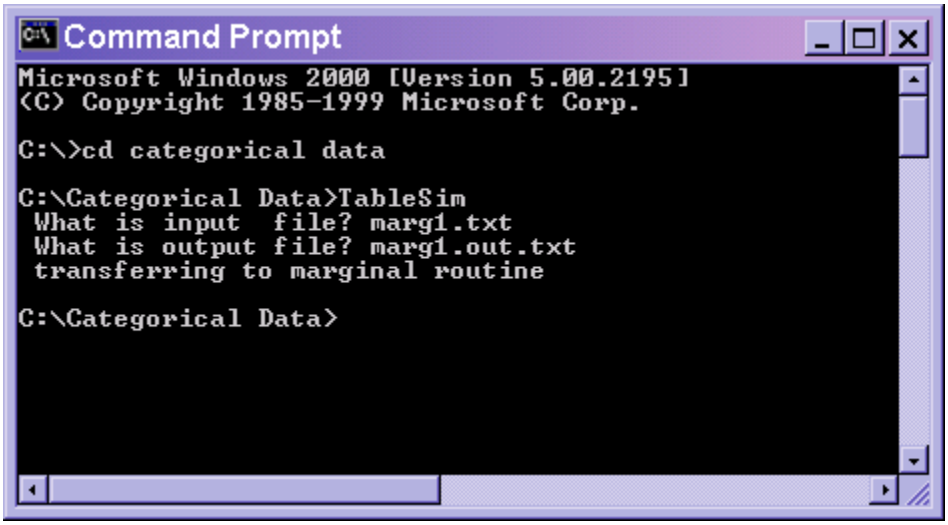

Figure 1a.—Command prompt execution of TableSim, a simple example. The program could have been invoked with only lowercase letters, i.e., as 'tablesim'. Note that the program tells the user which analysis it thinks it has been asked to run.

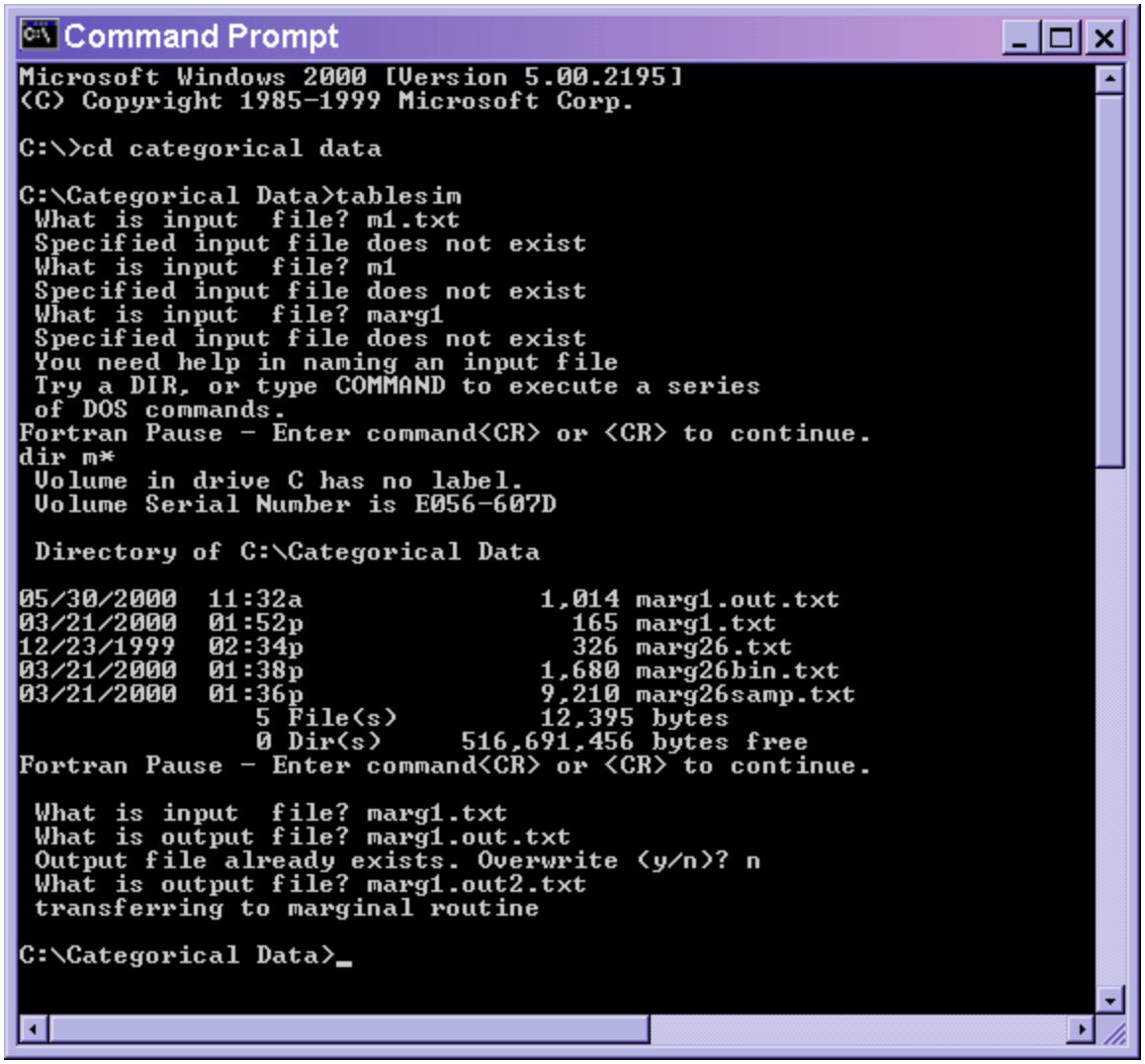

Figure 1b.—Command prompt execution of TableSim, a complex example. After three unsuccessful attempts to acquire an input file name, the program changes tactics. In this case, a simple 'dir m*' was used to list all files beginning with 'm' in the directory. Windows Explorer could also be used to figure out what the right file name is.

```
"Test of small sample marginal using 9x3 table (17.1)"
"marginal"
9 3
10000

0  1  0
8  1  8
0  1  0
0  1  0
0  1  0
0  1  0
0  1  0
1  0  1
1  0  1
```

Figure 2a.—Input file structure for analyzing independence in r x c tables. This example comes from table 17.1 of Mehta and Patel (1995).

- Line 1 is the analysis label; it must be ≤100 characters and enclosed in quotes.
- Line 2 is the keyword "marginal" (quotes required).
- Line 3 is the number of rows and columns.
- Line 4 is the number of random tables the program uses to estimate the P-value.
- Line 5 is blank.
- The next lines contain the observations; these must be integers.
- There is a blank line after the last row of observations.

```
Test of small sample marginal using 9x3 table (17.1)

Observed matrix:
     0     1     0
     8     1     8
     0     1     0
     0     1     0
     0     1     0
     0     1     0
     0     1     0
     1     0     1
     1     0     1

 Expected matrix:
     0.37    0.26    0.37
     6.30    4.41    6.30
     0.37    0.26    0.37
     0.37    0.26    0.37
     0.37    0.26    0.37
     0.37    0.26    0.37
     0.37    0.26    0.37
     0.74    0.52    0.74
     0.74    0.52    0.74

 Chi residual matrix:
    -0.61    1.45   -0.61
     0.68   -1.62    0.68
    -0.61    1.45   -0.61
    -0.61    1.45   -0.61
    -0.61    1.45   -0.61
    -0.61    1.45   -0.61
    -0.61    1.45   -0.61
     0.30   -0.72    0.30
     0.30   -0.72    0.30


For testing the table:
Based on  10000 samples, P(X^2 >=    22.099) = 0.02750
                        P(G^2 >=    23.297) = 0.03540
```

Figure 2b.—Output for analysis of independence of r x c tables; input file in figure 2a. (StatXact (Mehta and Patel 1995) reports exact $P(X^2) = 0.0269$; exact $P(G^2) = 0.0356$; asymptotic $P = 0.14$.)

```
"Test of small sample strata routine using 3x4x3 table"
"strata"
3 4 3
10000

2  3  1  2
0  3  0  0
4  1  1  1

3  0  1  3
0  4  3  0
4  3  0  3

0  1  4  1
4  1  3  4
4  1  3  3
```

Figure 3a.—Input file structure for assessing homogeneity of r x c table across k strata.

- Line 1 is the analysis label; it must be ≤100 characters and enclosed in quotes.
- Line 2 is the keyword "strata" (quotes required).
- Line 3 is the number of rows, the number of columns, and the number of strata.
- Line 4 is the number of random tables the program uses to estimate the P-value.
- Line 5 is blank.
- The next lines contain the observed values for the strata. The values must be integers; blank lines separate the strata.
- There is a blank line after the last row of observations.

Figure 3b.—Output for homogeneity of r x c table across k strata; input file in figure 3a. (StatXact reports Monte Carlo $P(X^2) = 0.043$, asymptotic $P(X^2) = 0.054$, Monte Carlo $P(G^2) = 0.035$, asymptotic $P(G^2) = 0.008$.)

```
Test of small sample strata routine using 3x4x3 table

Observed matrices
     2     3     1     2
     0     3     0     0
     4     1     1     1

     3     0     1     3
     0     4     3     0
     4     3     0     3

     0     1     4     1
     4     1     3     4
     4     1     3     3


 Expected matrices:
     1.27     1.01     1.52     1.52
     1.01     2.03     1.52     1.01
     3.04     1.27     1.01     1.77

     1.69     1.35     2.03     2.03
     1.35     2.70     2.03     1.35
     4.06     1.69     1.35     2.37

     2.04     1.63     2.45     2.45
     1.63     3.27     2.45     1.63
     4.90     2.04     1.63     2.86


 Chi residual matrices:
    0.65    1.97   -0.42    0.39
   -1.01    0.68   -1.23   -1.01
    0.55   -0.24   -0.01   -0.58


    1.01   -1.16   -0.72    0.68
   -1.16    0.79    0.68   -1.16
   -0.03    1.01   -1.16    0.41


   -1.43   -0.50    0.99   -0.93
    1.85   -1.25    0.35    1.85
   -0.41   -0.73    1.07    0.08


For testing the table:
Based on  10000 samples, P(X^2 >=    33.560) = 0.04330
                        P(G^2 >=    41.287) = 0.03560
```

This analysis would be useful in two main cases. The first is when the r x c structure is of interest, but is not modeled until any effect of the strata is determined. The second case is when the r x c structure is not of primary interest, but comparing that structure across strata is of interest. An example of the latter is r x c tables that are Markov chain transition count matrices, and the strata are different time periods. The analysis objective is to assess the homogeneity of the Markov chain structure.

Comparison of r x c table to theoretical distribution

This analysis compares an observed vector to an expected vector, an observed matrix to a common expected vector, or an observed matrix to an expected matrix. Figure 4a shows a sample input file; figure 4c shows the resulting output file. The theoretical distribution vectors are real numbers and, in principle, are the expected probabilities. Because the program always normalizes each expected vector to sum to 1.0, expected values or values proportional to the expected probabilities are also allowed. Figure 4b restates figure 4a using an alternative expected structure.

```
"Large sample test of theoretical distribution analysis"
'theory'
1 6
20000

45  60  70  54  80  64

1

.167  .167 .167  .167  .167  .167
```

Figure 4a.—Input file structure for comparing observed to theoretical distributions. As noted in text, quote marks can be single or double, but must match on a given line.

- Line 1 is the analysis label; it must be ≤ 100 characters and enclosed in quotes.
- Line 2 is the keyword 'theory' (quotes required).
- Line 3 is the number of rows and columns.
- Line 4 is the number of random tables the program uses to estimate the P-value.
- Line 5 is blank.
- The next lines contain the observed values; these must be integers.
- There is a blank line after the last row of observations.
- The next row contains the number of expected vectors; this number must be either '1' or the number of rows.
- The next line is blank.
- The next lines contain the theoretical distribution vectors. These are real numbers and, in principle, are the expected probabilities. Because the program always normalizes each expected vector to sum to '1', expected values or values proportional to the expected probabilities are also allowable input.
- There is a blank line after the last expected vector.

Figure 4b.—Input file structure for comparing observed to theoretical distributions using an alternative expression of the expected distribution. This formulation gives the same result as the formulation in figure 4a, despite using expected values that are not probabilities and improperly using integer instead of real values (no decimal points).

```
"Large sample test of theoretical distribution analysis"
'theory'
1 6
20000

45  60  70  54  80  64

1

1  1  1  1  1  1
```

Figure 4c.—Display of output file for comparison of observed to theoretical distribution; output corresponds to input file in figure 4a. (With 5 degrees of freedom, asymptotic $P(X^2)$ = 0.03416 and $P(G^2)$ = 0.03296.)

```
Large sample test of theoretical distribution analysis

Observed matrix:
    45    60    70    54    80    64

 Expected matrix:
    62.2   62.2   62.2   62.2   62.2   62.2

 Chi residual matrix:
    -2.18   -0.27    0.99   -1.04    2.26    0.23


For testing the table:
Based on  20000 samples, P(X^2 >=    12.046) = 0.03475
                         P(G^2 >=    12.137) = 0.03445
```

If the number of expected vectors is 1, then that vector is applied to either the single observed vector or to each row in the observed matrix, whichever case is applicable. An example of the latter case is a matrix with bird species as rows and habitat types as columns. The counts are number of birds of species *i* caught in habitat *j*. The common expected probability vector is the proportion of total net hours in each habitat. This analysis yields an omnibus test for random use of the landscape by the collection of bird species trapped.

Independence in r x c tables with repeated measures in the rows

Background—In the standard contingency table setup, only one observation is made on each independent experimental unit. It is not unusual, however, to have studies in which multiple observations are made on each experimental unit. Most often, this would take the form of multiple observations on individuals. For example, if we were studying differences in food preferences between males and females (populations), we would observe a number of independent organisms (clusters within populations). Each organism's preference distribution (response) would be ascertained from counting preferences for a number of trials. These preference counts would not be independent, and simply pooling results across individuals to assess sex-based differences would generally yield a misleading analysis.

The repeated measures can also take the form of what is often called aggregated data. In this case, a single observation is taken on each of a number of individuals in a group, but the aggregated observations don't follow the standard multinomial distribution. The repeated measures structure comes from viewing the groups as the sampling units, which are measured multiple times. For example, when studying spatial differences in species distributions we might define a set of areas (populations) and then randomly select a number of places to collect samples (clusters within populations). Each place's species composition (response) would be the usual frequency count of species present in the sample. Traditionally, samples within an area are pooled, and the resulting area x species table is analyzed for independence. Again, this will generally yield a misleading result. A more complete introduction to the aggregation problem in ecology can be found in Garson and Moser (1995).

There is no standard approach to analyzing this type of data set. The program implements three approaches and automatically runs all three. In the descriptions below, Population is indexed by j, and ranges from j = 1, …, J. Cluster is indexed by k, and ranges from k = 1, …, K_j. Response category is indexed by i, and ranges from i = 1, …, I. The number of responses in a particular cluster is given by n_{jk}.

Simple randomization—In the first approach, the test statistic is calculated by pooling across clusters and doing the standard test on the matrix of populations x responses. The test statistic is compared to a randomization distribution. Each population is allocated the appropriate number of clusters from the collection of all clusters using sampling without replacement, because under the null hypothesis all clusters come from the same underlying population. The resulting table is then analyzed with standard methods as described above. Repeated a sufficient number of times, this procedure generates a distribution for the test statistic that accounts for the correlated responses. The P-value is, as usual, based on the number of times the randomization statistic is equal to or larger than the test statistic. This analysis, because it shuffles a limited number of clusters, tends to be rather fast.

Manly's model—The second approach uses a model by Manly (1997, section 12.4). A standard chi-square test is computed for each population. The test statistic is the sum of the population statistics. As in the first approach, clusters are randomly allocated to populations. Within-population chi-square tests are computed and summed across populations to give a randomization value. Repeated many times, this procedure generates a distribution for the test statistic that accounts for the correlated responses. For this approach, the P-value is based on the number of times the randomization statistics are

equal to or smaller than the test statistic. The idea is that if the populations differ, then the within-population variation, as described by the within-population chi-square test, will be larger for the randomized arrangements of clusters than for the observed arrangement. Because it shuffles a limited number of clusters, this analysis also tends to be rather fast. The program's implementation is particularly fast because the cluster randomizations generated by the first analysis are re-used for this analysis.

Dirichlet-Multinomial Model

The third approach is a model-based analysis developed by Koehler and Wilson (1986, section 3), hereafter referred to as KW. A simple version of this analysis was presented in Garson and Moser (1995), hereafter referred to as GM. The analysis is based on the idea of there being more variability across clusters than is accounted for in the multinomial model used to derive the standard test. To save the non-statistician from translating KW, let's set up the procedure here.

Recall the basic structure of the problem. There are populations of interest (indexed by j, ranging in value from 1 to J), each of which has some number of clusters sampled (indexed by k, ranging in value from 1 to K_j). Individual observations on the clusters are classed into response categories (indexed by i, ranging from 1 to I). For a particular cluster in a given population, the true probability distribution for the response categories is given by the vector p_{jk}. KW treats this distribution as a random value from a Dirichlet distribution with mean vector $\boldsymbol{\pi}_j$ (the true mean vector of category probabilities for population j) and scale parameter γ_j (a measure of variability). The method estimates π_{ij} as n_{ij}/n_j, and p_{ijk} as n_{ijk}/n_{jk}, and uses these values to estimate the C_j parameter described in the KW and GM papers. C_j ranges from 1 (independent data) to n_j (very dispersed data). C_j is then used to estimate γ_j, which ranges from 0 (when $C_j = n_j$) to infinity (when $C_j = 1$). In this context, the scale parameter can be thought of as the "effective sample size"—when the scale parameter is small, the response probability distribution is not well determined and can vary a lot across clusters; when the scale parameter is at its maximum, there is no extra variability, so the cluster response distribution is always the population response distribution.

$\{C_j\}$ are also used to compute weights, $\{\alpha_j\}$. The weights are used to properly pool the population response distributions into an overall average distribution, $\boldsymbol{\pi}$. To generate the null distribution of the test statistics, the program sets $\boldsymbol{\pi}$, rather than $\boldsymbol{\pi}_j$, as the center of the Dirichlet distribution, while the scale parameters remain population-specific. Random response distributions for each population's clusters are created, and then an appropriate number of observations are generated for each cluster using those distributions. The test statistics take the form of X^2_{DM} in KW (equation 3.3).

The program uses the $\hat{C}_{uj}$ estimator for estimating C_j and thence γ_j. This is a more general version of the approach described in GM. Although Garson and Moser do not point it out, the GM procedure requires identical sample sizes in each cluster from a given population, i.e., $n_{jk} = n_j$. The $\hat{C}_{uj}$ estimator allows these sample sizes to vary in an arbitrary manner.

Sometimes, the experimental design suggests a repeated measures analysis, but the data that are collected appear to be independent. A nice feature of the KW approach is that if none of the sampled populations displayed overdispersion, the analysis adjusts gracefully and provides the same results as if running the small-sample test of independence. That this has happened is clear from C_j estimates equal to 1. This is much simpler than the equivalent ANOVA analysis, which requires determining if the repeated measures structure can be ignored and then running a separate ANOVA.

TableSim's implementation of the method has a few special features stemming from being a small-sample implementation of a large-sample method. First, it treats estimates of C_j below 1.05 as being equal to 1.00. Second, it perturbs cell counts away from zero when zeros cause computational problems. Third, this analysis can take a lot of time if there are many observations in the clusters. Because of this, the program prints a message to the screen after completing every 1,000 simulations. Fourth, when a population has only 1 cluster, the program uses an *ad hoc* averaging approach to guess a value for C_j. This averaging approach does not always work well—it can give values that are outside the allowed range or

values that are very different from those for the other populations and therefore suspect. So, the program displays the multi-cluster C_j values, and allows the user to alter the values assigned to single-cluster populations. In these circumstances, multiple runs can be made using different values for the assigned C_j values, thereby evaluating the impact on the P-value for the test of homogeneity. Sometimes, a single-cluster C_j value will generate the error message "Gamma < 0!" In such cases, a C_j value inside the estimated valid range printed by the program can be tried. While C_j values outside the valid range are guaranteed not to work, a sub-range inside the valid range may not work either for a single-cluster population. If the next try doesn't work either, the direction of movement of the printed γ_j value should be examined. The objective is to choose a C_j value that will make γ_j greater than zero.

File structure information—Figure 5a shows the input file structure for the problem presented in example 12.3 of Manly (1997). The populations are habitat types, the clusters within populations are snail colonies, and the response is shell type. The counts are number of snails in a colony with a particular shell type. The analysis asks whether shell type distribution varies across habitat types, while allowing colonies to have their own shell type distribution within each habitat type.

```
"Test of randomization using Manly shell data"
"randomize"
10  6
2  3  5  4  2  1
20000

15  24    0   0   76   39   0    0   1   1
17  25    0   0   41    7   0    0  57   9

 1   4    1   1    5   22   4    1  30  17
 0  24    1   0   51   80   7   16   0   6
 9   6    4   3  135  165  62  138  59   0

 2   1    3   0   33    4  11    4   8   0
 1   7   33  28   15   15  54   40  24   6
 0   0    4   2    6    5   9    5   4   0
 1   4   18  14   17    3   8   10  23   1
 7   0    3   4    7    0  12   11   8   0

 2   1   10   3    0    0   2    2   0   1
17  25   46  26    7   27  59   61   9   2
 5   2    5   6    4    1   5    3   0   0
 1   1   10   2    0    1  42   19   0   0

 0  66  185  23    0   32  82   11   8  20
 0  11   21  34   19    6  33   31   0   0

47  21    0   6    5   13   0    7  49  16
```

Figure 5a.—Input file structure for analyzing independence in r x c tables with repeated measures. The Manly shell data problem is described in the text.

- Line 1 is the analysis label; it is ≤ 100 characters, and must be enclosed in quotes.
- Line 2 is the keyword "randomize" (quotes required).
- Line 3 is the number of response categories (columns) and populations (rows).
- Line 4 lists the number of clusters (sub-rows) in each of the populations.
- Line 5 is the number of random tables the program uses to estimate the P-value.
- Line 6 is blank.
- The next lines contain the observed values; these must be integers.
- There is a blank line after each population of clusters, including the last.

Figure 5b shows the output file for the analysis. For this analysis, the output file has three parts. The first part is the standard restatement of the analysis problem. The second part presents the Pearson X^2 and likelihood ratio G^2 statistics and associated P-values for the simple randomization and Manly analyses. The third part presents the KW analysis. The KW output begins with the cluster-specific response probability distributions; these tables give some sense of the intrapopulation variability relevant to the KW analysis. A parameter estimate table follows, displaying the C_j, γ_j, and α_j values. A γ_j value of –99.000 indicates that the null response distribution was used as the actual response distribution for all clusters in that population. This will generally be associated with a C_j value of 1. The response category probabilities are displayed for each population, along with the weighted null distribution. The unweighted null distribution is also printed to show how much change was caused by violations of assumptions in the data set structure. The output file concludes with the KW test statistics and associated P-values.

Figure 5b.—Display of output file for the analysis of independence of r x c tables with repeated measures; output based on input file in figure 5a. The C_j value of 31.00 for population 6 was assigned. Manly (1997) reported a randomization probability of 0.001 using his method (n = 1000 samples).

```
Test of randomization using Manly shell data
Observed matrices
    15    24     0     0    76    39     0     0     1     1
    17    25     0     0    41     7     0     0    57     9

     1     4     1     1     5    22     4     1    30    17
     0    24     1     0    51    80     7    16     0     6
     9     6     4     3   135   165    62   138    59     0

     2     1     3     0    33     4    11     4     8     0
     1     7    33    28    15    15    54    40    24     6
     0     0     4     2     6     5     9     5     4     0
     1     4    18    14    17     3     8    10    23     1
     7     0     3     4     7     0    12    11     8     0

     2     1    10     3     0     0     2     2     0     1
    17    25    46    26     7    27    59    61     9     2
     5     2     5     6     4     1     5     3     0     0
     1     1    10     2     0     1    42    19     0     0

     0    66   185    23     0    32    82    11     8    20
     0    11    21    34    19     6    33    31     0     0

    47    21     0     6     5    13     0     7    49    16


Collapsed observed matrix:
    32   49    0    0  117   46    0    0   58   10
    10   34    6    4  191  267   73  155   89   23
    11   12   61   48   78   27   94   70   67    7
    25   29   71   37   11   29  108   85    9    3
     0   77  206   57   19   38  115   42    8   20
    47   21    0    6    5   13    0    7   49   16


Under simple randomization:
Based on  20000 samples, P(X^2 >  1829.912) = 0.00080
                        P(G^2 >  1922.039) = 0.00020

Under the Manly test:
Based on  20000 samples, P(X^2 <   824.814) = 0.00025
                        P(G^2 <   803.702) = 0.00020

K-W analysis output
Cluster-specific response probabilities
Pop   Cluster          Response Probability Distribution
  1      1      0.0962  0.1538  0.0000  0.0000  0.4872  0.2500  0.0000  0.0000  0.0064  0.0064
  1      2      0.1090  0.1603  0.0000  0.0000  0.2628  0.0449  0.0000  0.0000  0.3654  0.0577

  2      1      0.0116  0.0465  0.0116  0.0116  0.0581  0.2558  0.0465  0.0116  0.3488  0.1977
  2      2      0.0000  0.1297  0.0054  0.0000  0.2757  0.4324  0.0378  0.0865  0.0000  0.0324
  2      3      0.0155  0.0103  0.0069  0.0052  0.2324  0.2840  0.1067  0.2375  0.1015  0.0000

  3      1      0.0303  0.0152  0.0455  0.0000  0.5000  0.0606  0.1667  0.0606  0.1212  0.0000
```

Figure 5b.—continued

```
  3      2      0.0045  0.0314  0.1480  0.1256  0.0673  0.0673  0.2422  0.1794  0.1076  0.0269
  3      3      0.0000  0.0000  0.1143  0.0571  0.1714  0.1429  0.2571  0.1429  0.1143  0.0000
  3      4      0.0101  0.0404  0.1818  0.1414  0.1717  0.0303  0.0808  0.1010  0.2323  0.0101
  3      5      0.1346  0.0000  0.0577  0.0769  0.1346  0.0000  0.2308  0.2115  0.1538  0.0000

  4      1      0.0952  0.0476  0.4762  0.1429  0.0000  0.0000  0.0952  0.0952  0.0000  0.0476
  4      2      0.0609  0.0896  0.1649  0.0932  0.0251  0.0968  0.2115  0.2186  0.0323  0.0072
  4      3      0.1613  0.0645  0.1613  0.1935  0.1290  0.0323  0.1613  0.0968  0.0000  0.0000
  4      4      0.0132  0.0132  0.1316  0.0263  0.0000  0.0132  0.5526  0.2500  0.0000  0.0000

  5      1      0.0000  0.1546  0.4333  0.0539  0.0000  0.0749  0.1920  0.0258  0.0187  0.0468
  5      2      0.0000  0.0710  0.1355  0.2194  0.1226  0.0387  0.2129  0.2000  0.0000  0.0000

  6      1      0.2866  0.1280  0.0000  0.0366  0.0305  0.0793  0.0000  0.0427  0.2988  0.0976

Pop   # clusters    C(j)     gamma(j)   alpha(j)
  1        2       10.372     15.539      0.157
  2        3       35.277     11.955      0.126
  3        5        6.665     24.024      0.371
  4        4        9.602     23.168      0.221
  5        2       30.593     10.947      0.099
  6        1       31.000      4.433      0.028

Population                   Response probabilities
   1        0.1026  0.1571  0.0000  0.0000  0.3750  0.1474  0.0000  0.0000  0.1859  0.0321
   2        0.0117  0.0399  0.0070  0.0047  0.2242  0.3134  0.0857  0.1819  0.1045  0.0270
   3        0.0232  0.0253  0.1284  0.1011  0.1642  0.0568  0.1979  0.1474  0.1411  0.0147
   4        0.0614  0.0713  0.1744  0.0909  0.0270  0.0713  0.2654  0.2088  0.0221  0.0074
   5        0.0000  0.1323  0.3540  0.0979  0.0326  0.0653  0.1976  0.0722  0.0137  0.0344
   6        0.2866  0.1280  0.0000  0.0366  0.0305  0.0793  0.0000  0.0427  0.2988  0.0976
wtd null    0.0475  0.0713  0.1220  0.0688  0.1578  0.1079  0.1622  0.1319  0.1090  0.0216

unwtd null  0.0448  0.0795  0.1232  0.0544  0.1508  0.1504  0.1397  0.1286  0.1003  0.0283


Under the KW model:
Based on  20000 samples, P(X^2 >=    96.655) = 0.00080
                         P(G^2 >=   108.512) = 0.00005
```

Benchmarking—It is difficult to check the performance of this routine, because there are no benchmarks for comparison. However, one fundamental check for the first analysis method is how it behaves when each individual in a group has only one observation – this is the underlying structure of the standard test of independence in r x c tables, for which checks are available. Figure 6a provides a constructed data set for this check; figure 6b provides the results. The close agreement between the "randomization" routine results and the "marginal" routine results suggests that the software behaves as desired. In a similar vein, for the KW analysis, we can check the behavior when $\{C_j\} = 1$. The analysis should give the results similar to the "marginal" routine, and it does (fig. 7a, b). We can also create a large-sample KW analysis, and compare the program's output to the expected large-sample results. The results (fig. 8a, b) show the program behaving properly.

```
'Test of randomization using single observations per individual'
'randomize'
3 3
31 14 22
10000

1 0 0                   0 1 0
1 0 0                   0 1 0                   1 0 0
1 0 0                   0 1 0                   1 0 0
1 0 0                   0 1 0                   1 0 0
1 0 0                   0 1 0                   1 0 0
1 0 0                   0 1 0                   1 0 0
1 0 0                   0 0 1                   0 1 0
1 0 0                   0 0 1                   0 1 0
1 0 0                                           0 1 0
1 0 0                   1 0 0                   0 1 0
1 0 0                   1 0 0                   0 1 0
1 0 0                   1 0 0                   0 1 0
1 0 0                   0 1 0                   0 1 0
1 0 0                   0 1 0                   0 1 0
1 0 0                   0 1 0                   0 1 0
1 0 0                   0 1 0                   0 1 0
1 0 0                   0 1 0                   0 0 1
1 0 0                   0 1 0                   0 0 1
1 0 0                   0 1 0                   0 0 1
1 0 0                   0 0 1                   0 0 1
1 0 0                   0 0 1                   0 0 1
1 0 0                   0 0 1                   0 0 1
0 1 0                   0 0 1                   0 0 1
0 1 0                   0 0 1                   0 0 1
```

Figure 6a.—Input file structure for checking the simple randomization routine against the marginal routine using a constructed data set. Because of the length of this data set, the observations have been strung into columns; in the actual data file, there is only one column of data.

```
Test of randomization using single observations per individual

Collapsed observed matrix:
    21    8    2
     3    6    5
     5    9    8

Under simple randomization:
Based on  10000 samples, P(X^2 >=    15.754) = 0.00280
                         P(G^2 >=    16.861) = 0.00390


Test of randomization using single observations per individual; marginal

Observed matrix:
    21     8     2
     3     6     5
     5     9     8

 Expected matrix:
    13.42   10.64    6.94
     6.06    4.81    3.13
     9.52    7.55    4.93

 Chi residual matrix:
     2.07   -0.81   -1.88
    -1.24    0.54    1.05
    -1.47    0.53    1.39


For testing the table:
Based on  10000 samples, P(X^2 >=    15.754) = 0.00200
                         P(G^2 >=    16.861) = 0.00280
```

Figure 6b.—The first section summarizes relevant output from the input file in figure 6a. The second section displays the output file for the equivalent marginal analysis. For the marginal analysis, StatXact reports exact $P(X^2) = 0.0028$ and $P(G^2) = 0.0035$.

```
"Randomization test of independent data, Cleth. gapperi captures"
"randomize"
4  2
3  3
20000

  2    1     0    3
  0    4     2    4
  2    4     1    3

  0    0     0    2
  0    0     1    0
  0    0     2    1


"Marginal test of independent data, Cleth. gapperi captures"
"marginal"
2  4
20000

  4    9     3   10
  0    0     3    3
```

Figure 7a.—Input file structure for checking KW randomization routine against marginal routine using a data set with $\{C_j\} = 1$.

Randomization test of independent data in repeated measures design

```
Collapsed observed matrix:
     4    9    3   10
     0    0    3    3

Under simple randomization:
Based on  20000 samples, P(X^2 >=     7.006) = 0.20110
                         P(G^2 >=     8.522) = 0.20110

Under the Manly test:
Based on  20000 samples, P(X^2 <=     8.726) = 0.09285
                         P(G^2 <=    11.804) = 0.20110

K-W analysis output
Cluster-specific response probabilities
Pop   Cluster          Response Probability Distribution
  1      1      0.3333  0.1667  0.0000  0.5000
  1      2      0.0000  0.4000  0.2000  0.4000
  1      3      0.2000  0.4000  0.1000  0.3000

  2      1      0.0000  0.0000  0.0000  1.0000
  2      2      0.0000  0.0000  1.0000  0.0000
  2      3      0.0000  0.0000  0.6667  0.3333

Pop   # clusters    C(j)     gamma(j)   alpha(j)
  1        3        1.000    -99.000      0.813
  2        3        1.000    -99.000      0.188

Population                   Response probabilities
   1        0.1538  0.3462  0.1154  0.3846
   2        0.0000  0.0000  0.5000  0.5000
wtd null    0.1250  0.2813  0.1875  0.4063

unwtd null  0.1250  0.2813  0.1875  0.4063


Under the KW model:
Based on  20000 samples, P(X^2 >=     7.006) = 0.06400
                         P(G^2 >=     8.522) = 0.04570
```

Figure 7b.—The first section summarizes output from the first analysis of the input file in figure 7a; the second section summarizes the second analysis output.

Randomization test of independent data in repeated measures design; marginal

```
For testing the table:
Based on  20000 samples, P(X^2 >=     7.006) = 0.05185
                         P(G^2 >=     8.522) = 0.05570
```

```
"Test of asymptotic randomization using modified Fraser krill data"
"randomize"
8  5
9  8  9  8  8
10000

 20   70   521  594   154  441   87   55
 58   85   271  105   556  391  127    8
  2   47   169  236   716  633  144   15
 18   60    21  694   154  341   87   55
 58   85   171  205   556  491  127    8
  2   97    69  236   716  563  144    5
 22   50   121  994   154  341   87   55
 58   85   171  505   556  291  127   18
  2   17    69  236   716  363  144    5

 38   35    58  102   122  162  201    5
 52    2    62  867   265  134   89    7
 11  124   182  462   751  596  123   93
 32  100   320  263   309  660  229   49
  8   35    58  202   122  162  101    9
 12    2    62   67   265  134  189   10
 41  124   182  362   751  396  223   53
112  100   320  263   309  660   29   49

 40   32    62  215   517  975  344   19
  3   71    41  608   500  272   19   10
 42  172   730  658   680  515  219    2
 40   72   162  115   517  675  444   29
  3   71   641  708   712  172   19   15
142  272   230  958   380  215   19   12
 60   17   162  215   517  875  444  119
 13   71    41  708   500  172   19   20
 92  252   550 1058   380  315   19   12

 23  186   347  440   153  180  427   12
 39   18    71  582   813  244  189   50
 27  186   447  440   273  380   27   92
 22    8    71  582  1213  444  289   60
 53  186   447  440   253  580   27   22
 15    8    71  582   833  444   89   55
 23  186   447  440   253  880  227   32
 10    8    71  582   913  444   89   40

 11    8    30  941    99  250  624   20
 58   85   171  405   556  491  127    8
  2   17    69  236   716  763   44    5
 38   35    58  102   122  162   10   30
 52    2    62   67   265  134  289   10
 40  172   162  215   617  675  104  119
  3   71   641  708   500  172   19   22
 23  186   447  440    53   80   27   32
```

Figure 8a.—Input file structure for checking large-sample KW randomization routine against analytical large-sample results.

```
Test of asymptotic randomization using modified Fraser krill data

Collapsed observed matrix:
   240  596 1583 3805 4278 3855 1074  224
   306  522 1244 2588 2894 2904 1184  275
   435 1030 2619 5243 4703 4186 1546  238
   212  786 1972 4088 4704 3596 1364  363
   227  576 1640 3114 2928 2727 1244  246

K-W analysis output
Pop   # clusters    C(j)     gamma(j)   alpha(j)
  1        9        9.583    203.505      0.295
  2        8       10.943    176.249      0.197
  3        9       17.753    137.155      0.203
  4        8       15.055    153.663      0.205
  5        8       22.972     80.330      0.100

Population                  Response probabilities
   1        0.0153  0.0381  0.1011  0.2431  0.2733  0.2462  0.0686  0.0143
   2        0.0257  0.0438  0.1044  0.2172  0.2428  0.2437  0.0994  0.0231
   3        0.0217  0.0515  0.1310  0.2621  0.2351  0.2093  0.0773  0.0119
   4        0.0124  0.0460  0.1154  0.2393  0.2753  0.2105  0.0798  0.0212
   5        0.0179  0.0453  0.1291  0.2452  0.2305  0.2147  0.0979  0.0194
wtd null    0.0183  0.0443  0.1136  0.2413  0.2557  0.2277  0.0817  0.0175

unwtd null  0.0184  0.0454  0.1171  0.2435  0.2522  0.2232  0.0829  0.0174


Under the KW model:
Based on  10000 samples, P(X^2 >    52.224) = 0.00500
                         P(G^2 >    52.027) = 0.00410
```

Figure 8b.—Summary of output file for test of large-sample KW randomization analysis of independence of r x c tables with repeated measures; input file shown in figure 8a. Asymptotically, with 28 degrees of freedom, $P(X^2) = 0.0037$ and $P(G^2) = 0.0038$. Note that a naïve analysis would conclude that $X^2 \approx G^2 = 695$, with $P << 0.00001$.

A krill sampling example—Now that we understand what the repeated measures routine is doing, and have confidence that it is doing it correctly, let us work through an example in some detail. Consider comparing the size distribution of krill in the diets of five penguins in various years, as described in Fraser and Hofmann (in press) and Hofmann and Fraser (in press). In our abstract terms, the problem has five popula-tions (the penguins), a variable number of clusters within populations (the years, ranging from 1 to 4), and an 8-category response (size class counts of the krill taken from a penguin's stomach in a particular year; see figure 9a for data).

When we run this problem through the program, it asks us what to do about the value of C_j for the fifth population, which has only 1 cluster. One response is to let the program use the estimate it generated; the results of this decision are displayed in figure 9b. Another response is to tell the pro-gram to use a different value; the results of one such choice are displayed in figure 9c.

The output displayed in figure 9b shows that the test statistic values for the plain randomization methods are rather large, but the associated P-values are also rather large. In the KW output section, the $\{C_j\}$ are much larger than 1 and the $\{\gamma_j\}$ show that the effective sample sizes are much smaller than the organism counts would indicate. The differences between the unweighted null response distribution and the proper weighted null distribution show the effects of the $\{C_j\}$ being different. Compare this to the results in figure 8b, where the $\{C_j\}$ are clearly greater than 1, but are fairly similar among themselves. There are still clear effects on the tests, but the weighted and unweighted response distributions are much closer. Finally, when we look at the test statistics, it is clear that the KW algorithm reduced the test statistic values by two to three orders of magnitude. Such a large impact is not too surprising given the relation of the $\{\gamma_j\}$ to the observed counts.

Considering the output displayed in figure 9c, we see that for this data set a fairly large perturbation in C_5

```
"Test of randomization using Fraser krill data"
"randomize"
8  5
3  4  3  2  1
10000

  0    0    21   94   154  441   87   55
 58   85   171  405   556  491  127    8
  2   17    69  236   716  763  144    5

 38   35    58  102   122  162  201    0
 52    2    62   67   265  134  289    0
 61  124   182  362   751  696  423  223
 92  100   320  263   309  660  429   49

 40  172   162  215   517  675  444  119
  3   71   641  708   500  172   19    0
142  372  1330  858   380  115   19    2

 23  186   447  440   253   80   27    2
  0    8    71  582  1013  444   89    0

  0    1    30   41   199  250   24    0
```

Figure 9a.—Input file for krill example. Five populations, with from 1 to 4 clusters sampled per population. Size class counts are in the columns.

had relatively small effects on the weighted response distribution and moderate impacts on the test statistics. However, the estimated P-values were reduced by about 50 percent.

Having an explicit approach to how the data structure deviates from a standard multinomial creates a very firm foundation for generating the null distribution of the test statistics. Although the Manly shell data (figure 5) showed that the three analyses can give fairly consistent results for some data structures, it appears equally clear that the KW analysis can be much more powerful than the other two analyses. Moreover, the degree of similarity between the KW and plain randomization methods depends on the value chosen for C_6. The KW P-values are much smaller than the other P-values for some C_6 values. A strong argument can be made that results from the two plain randomization methods should be treated as being presented only for historical completeness. Indeed, these two methods are likely to be deleted in future versions of TableSim.

Multiple Analyses

A single input file can contain multiple analyses for the program to run. Simply leave one or more blank lines after the data for a given analysis before starting the control language for the next analysis. All results are written to the same output file. Figure 10a shows an example input file; figure 10b shows the corresponding output file. This feature is particularly useful on slower machines, because it allows the user to run a number of analyses on the system overnight. On faster machines, this feature allows the user to set up some analyses and let them run in the background while other tasks are performed in the foreground.

```
Test of randomization using Fraser krill data

Observed matrices
     0     0    21    94   154   441    87    55
    58    85   171   405   556   491   127     8
     2    17    69   236   716   763   144     5

    38    35    58   102   122   162   201     0
    52     2    62    67   265   134   289     0
    61   124   182   362   751   696   423   223
    92   100   320   263   309   660   429    49

    40   172   162   215   517   675   444   119
     3    71   641   708   500   172    19     0
   142   372  1330   858   380   115    19     2

    23   186   447   440   253    80    27     2
     0     8    71   582  1013   444    89     0

     0     1    30    41   199   250    24     0


Collapsed observed matrix:
    60  102  261  735 1426 1695  358   68
   243  261  622  794 1447 1652 1342  272
   185  615 2133 1781 1397  962  482  121
    23  194  518 1022 1266  524  116    2
     0    1   30   41  199  250   24    0


Under simple randomization:
Based on  10000 samples, P(X^2 >=  4766.887) = 0.23800
                         P(G^2 >=  4718.594) = 0.28750

Under the Manly test:
Based on  10000 samples, P(X^2 <=  5566.043) = 0.26760
                         P(G^2 <=  5826.359) = 0.28750

K-W analysis output
Cluster-specific response probabilities
Pop   Cluster          Response Probability Distribution
  1      1      0.0000  0.0000  0.0246  0.1103  0.1808  0.5176  0.1021  0.0646
  1      2      0.0305  0.0447  0.0900  0.2130  0.2925  0.2583  0.0668  0.0042
  1      3      0.0010  0.0087  0.0353  0.1209  0.3668  0.3909  0.0738  0.0026

  2      1      0.0529  0.0487  0.0808  0.1421  0.1699  0.2256  0.2799  0.0000
  2      2      0.0597  0.0023  0.0712  0.0769  0.3042  0.1538  0.3318  0.0000
  2      3      0.0216  0.0439  0.0645  0.1283  0.2661  0.2466  0.1499  0.0790
  2      4      0.0414  0.0450  0.1440  0.1184  0.1391  0.2970  0.1931  0.0221

  3      1      0.0171  0.0734  0.0691  0.0917  0.2206  0.2880  0.1894  0.0508
  3      2      0.0014  0.0336  0.3032  0.3349  0.2365  0.0814  0.0090  0.0000
  3      3      0.0441  0.1156  0.4133  0.2666  0.1181  0.0357  0.0059  0.0006

  4      1      0.0158  0.1276  0.3066  0.3018  0.1735  0.0549  0.0185  0.0014
  4      2      0.0000  0.0036  0.0322  0.2637  0.4590  0.2012  0.0403  0.0000

  5      1      0.0000  0.0018  0.0550  0.0752  0.3651  0.4587  0.0440  0.0000

Pop   # clusters    C(j)     gamma(j)   alpha(j)
  1        3       55.878     30.547      0.292
  2        4       45.974     46.495      0.501
  3        3      232.521     10.429      0.115
  4        2      173.368     10.070      0.073
  5        1      103.839      4.290      0.018

Population                  Response probabilities
   1        0.0128  0.0217  0.0555  0.1562  0.3031  0.3603  0.0761  0.0145
   2        0.0366  0.0393  0.0938  0.1197  0.2182  0.2491  0.2023  0.0410
   3        0.0241  0.0801  0.2779  0.2320  0.1820  0.1253  0.0628  0.0158
   4        0.0063  0.0529  0.1413  0.2789  0.3454  0.1430  0.0317  0.0005
   5        0.0000  0.0018  0.0550  0.0752  0.3651  0.4587  0.0440  0.0000
wtd null    0.0253  0.0392  0.1065  0.1541  0.2509  0.2634  0.1340  0.0266

unwtd null  0.0220  0.0505  0.1535  0.1883  0.2469  0.2189  0.1000  0.0199


Under the KW model:
Based on  10000 samples, P(X^2 >=    43.941) = 0.04880
                         P(G^2 >=    43.170) = 0.00940
```

Figure 9b.—Output file for test of krill data; input file shown in figure 9a. C(j) proposed by program for population 5 accepted by user. Note that a naïve analysis would conclude $X^2 = 4767$ and $G^2 = 4719$, and with 28 degrees of freedom $P << 0.00001$.

Figure 9c.—Summary of output file for second test of krill data; input file shown in figure 9a. C(j) proposed by program for population 5 was changed by user to “fit in” with lower value populations. This change affected only the KW analysis output.

```
Test of randomization using Fraser krill data

K-W analysis output
Cluster-specific response probablities
Pop   Cluster          Response Probability Distribution
  1      1      0.0000  0.0000  0.0246  0.1103  0.1808  0.5176  0.1021  0.0646
  1      2      0.0305  0.0447  0.0900  0.2130  0.2925  0.2583  0.0668  0.0042
  1      3      0.0010  0.0087  0.0353  0.1209  0.3668  0.3909  0.0738  0.0026

  2      1      0.0529  0.0487  0.0808  0.1421  0.1699  0.2256  0.2799  0.0000
  2      2      0.0597  0.0023  0.0712  0.0769  0.3042  0.1538  0.3318  0.0000
  2      3      0.0216  0.0439  0.0645  0.1283  0.2661  0.2466  0.1499  0.0790
  2      4      0.0414  0.0450  0.1440  0.1184  0.1391  0.2970  0.1931  0.0221

  3      1      0.0171  0.0734  0.0691  0.0917  0.2206  0.2880  0.1894  0.0508
  3      2      0.0014  0.0336  0.3032  0.3349  0.2365  0.0814  0.0090  0.0000
  3      3      0.0441  0.1156  0.4133  0.2666  0.1181  0.0357  0.0059  0.0006

  4      1      0.0158  0.1276  0.3066  0.3018  0.1735  0.0549  0.0185  0.0014
  4      2      0.0000  0.0036  0.0322  0.2637  0.4590  0.2012  0.0403  0.0000

  5      1      0.0000  0.0018  0.0550  0.0752  0.3651  0.4587  0.0440  0.0000

Pop   # clusters    C(j)     gamma(j)   alpha(j)
  1        3       55.878     30.547      0.287
  2        4       45.974     46.495      0.492
  3        3      232.521     10.429      0.112
  4        2      173.368     10.070      0.072
  5        1       50.000     10.102      0.037

Population                    Response probabilities
   1        0.0128  0.0217  0.0555  0.1562  0.3031  0.3603  0.0761  0.0145
   2        0.0366  0.0393  0.0938  0.1197  0.2182  0.2491  0.2023  0.0410
   3        0.0241  0.0801  0.2779  0.2320  0.1820  0.1253  0.0628  0.0158
   4        0.0063  0.0529  0.1413  0.2789  0.3454  0.1430  0.0317  0.0005
   5        0.0000  0.0018  0.0550  0.0752  0.3651  0.4587  0.0440  0.0000
wtd null    0.0248  0.0385  0.1055  0.1526  0.2531  0.2672  0.1322  0.0261

unwtd null  0.0220  0.0505  0.1535  0.1883  0.2469  0.2189  0.1000  0.0199


Under the KW model:
Based on  10000 samples, P(X^2 >=    46.502) = 0.02240
                         P(G^2 >=    45.912) = 0.00520
```

```
"Test of small sample marginal using 9x3 table (17.1)"
"marginal"
9 3
1000

0  1  0
8  1  8
0  1  0
0  1  0
0  1  0
0  1  0
0  1  0
1  0  1
1  0  1



"Large sample test of theoretical distribution analysis"
'theory'
1 6
1000

45  60  70  54  80  64

1

.167  .167 .167  .167  .167  .167
```

Figure 10a.—Example input file structure for performing multiple analyses in a single run.

Figure 10b.—Example output file structure for performing multiple analyses in a single run; output corresponds to input file shown in figure 10a.

```
Test of small sample marginal using 9x3 table (17.1)

Observed matrix:
     0     1     0
     8     1     8
     0     1     0
     0     1     0
     0     1     0
     0     1     0
     0     1     0
     1     0     1
     1     0     1

 Expected matrix:
     0.37    0.26    0.37
     6.30    4.41    6.30
     0.37    0.26    0.37
     0.37    0.26    0.37
     0.37    0.26    0.37
     0.37    0.26    0.37
     0.37    0.26    0.37
     0.74    0.52    0.74
     0.74    0.52    0.74

 Chi residual matrix:
    -0.61    1.45   -0.61
     0.68   -1.62    0.68
    -0.61    1.45   -0.61
    -0.61    1.45   -0.61
    -0.61    1.45   -0.61
    -0.61    1.45   -0.61
    -0.61    1.45   -0.61
     0.30   -0.72    0.30
     0.30   -0.72    0.30


For testing the table:
Based on   1000 samples, P(X^2 >=    22.099) = 0.02100
                         P(G^2 >=    23.297) = 0.02400


Large sample test of theoretical distribution analysis

Observed matrix:
    45    60    70    54    80    64

 Expected matrix:
    62.2   62.2   62.2   62.2   62.2   62.2

 Chi residual matrix:
    -2.18   -0.27    0.99   -1.04    2.26    0.23


For testing the table:
Based on   1000 samples, P(X^2 >=    12.046) = 0.04100
                         P(G^2 >=    12.137) = 0.03700
```

User Notes

1. Open a command window via “Start → Programs → MS-DOS Prompt” in Windows 9x, “Start → Programs → Command Prompt” in Windows NT, and “Start → Programs → Accessories → Command Prompt” in Windows 2000. Alternatives include starting TableSim from the “Run” item in the “Start” menu or double-clicking on the program's entry in Windows Explorer.
2. Do not use spaces in file names when working in Windows 9x or Windows NT. However, in Windows 2000 and Windows XP, spaces in file names are allowed.
3. When you finish using TableSim, type ‘exit’ at the prompt to close the command window.

Source Code Issues

The program was written in the Compaq® Visual Fortran environment, compiler version 6.1a (DEC 1997), running under Microsoft® Windows® 2000 Professional. The source code was compiled with switches at default settings. The fixed format coding style reflects the origin of these routines, but all code is compliant with the Fortran 90 standard and has generally been updated to reflect F90 rather than F77 programming practice. The use of a command line interface simplifies programming in Microsoft environments and makes porting the software to other operating systems much easier.

There are three main reasons to modify the source code. The first is to alter the allowed line length of files. This requires modifying the RECL specifier in the three OPEN statements in subroutine FILENAM. For example, OPEN(junit, FILE=fname, RECL=180) would allow lines of up to 180 characters to be read or written, depending on which of the OPEN statements was altered. As this is standard Fortran, this modification is relevant for all compilers.

The second reason is to run the program using an executable created by a different compiler. The uniform random number generator, RANU, uses the Compaq library DFLIB. This library has a non-standard subroutine GETTIM, which is called when initializing the generator. Therefore, the random number generator's initialization routine would need to be rewritten to get the program to compile and execute correctly.

The third reason is simply to extend the program's capabilities – handling higher order tables, ordered data, Bayesian analysis, and so on.

Literature Cited

Agresti, Alan. 1990.
Categorical data analysis. New York, NY: John Wiley & Sons. 558 p.

Agresti, Alan. 1996.
An introduction to categorical data analysis. New York, NY: John Wiley & Sons. 290 p.

DEC. 1997.
DIGITAL Fortran language reference manual. Maynard, MA: Digital Equipment Corporation.

Fienberg, Stephen E. 1980.
The analysis of cross-classified categorical data. 2d ed. Cambridge, MA: The MIT Press. 198 p.

Fraser, William R.; Hofmann, E.E. In press.
Krill-sea ice interactions, part I: A predator's perspective on causal links between climate change, physical forcing and ecosystem response. Marine Ecology Progress Series.

Garson, Glenn I.; Moser, E. Barry. 1995.
Aggregation and the Pearson chi-square statistic for homogeneous proportions and distributions in ecology. Ecology. 76: 2258-2269.

Hofmann, E.E.; Fraser, William R. In press.
Krill-sea ice interactions, part II: A coupled ecological-environmental model. Marine Ecology Progress Series.

Koehler, Kenneth J.; Wilson, Jeffrey R. 1986.
Chi-square tests for comparing vectors of proportions for several cluster samples. Communications in Statistics: part A. Theory and Methods. 15: 2977-2990.

Manly, Bryan F. J. 1997.
Randomization, bootstrap and monte carlo methods in biology. 2d ed. London: Chapman & Hall. 399 p.

Mehta, Cyrus; Patel, Nitin. 1995.
StatXact 3 for Windows: Statistical Software for Exact Nonparametric Inference. User Manual. Cambridge, MA: Cytel Corporation. 788 p.

Rugg, David J.
2003. **TableSim—A program for analysis of small-sample categorical data**. Gen. Tech. Rep. NC-232. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station. 27 p.

Documents a computer program for calculating correct P-values of 1-way and 2-way tables when sample sizes are small. The program is written in Fortran 90; the executable code runs in 32-bit Microsoft® command line environments.

KEY WORDS: Fortran, aggregated data, randomization, Dirichlet-multinomial.

MISSION STATEMENT

We believe the good life has its roots in clean air, sparkling water, rich soil, healthy economies and a diverse living landscape. Maintaining the good life for generations to come begins with everyday choices about natural resources. The North Central Research Station provides the knowledge and the tools to help people make informed choices. That's how the science we do enhances the quality of people's lives.

For further information contact:

North Central
Research Station
USDA Forest Service
1992 Folwell Ave., St. Paul, MN 55108

Or visit our web site:
www.ncrs.fs.fed.us