APPLE-VECTA DOCUMENTATION

Phyllis A. Gimotty

September 1982

Report Number UMTRI-82-51

University of Michigan
TRANSPORTATION RESEARCH INSTITUTE
Ann Arbor, Michigan 48109

·

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
UMTRI-82-51		

4. Title and Subtitle	5. Report Date
APPLE-VECTA Documentation	September 1982
	6. Performing Organization Code

7. Author's)	8. Performing Organization Report No.
Phyllis A. Gimotty	UMTRI-82-51

9. Performing Organization Name and Address	10. Work Unit No.
University of Michigan Transportation Research Institute Ann Arbor, Michigan 48109	11. Contract or Grant No. Grant No. 5-A19
	13. Type of Report and Period Covered

12. Sponsoring Agency Name and Address	13. Type of Report and Period Covered
National Highway Traffic Safety Admin. Department of Transportation Washington, D.C. 20590	Final Sept. 1981-Sept. 1982
	14. Sponsoring Agency Code

15. Supplementary Notes

16. APPLE-VECTA is a microcomputer program adapted from VECTA, a model originally developed by J.E.K. Smith for conducting iterative proportional fitting during analyses of contingency tables. With the limited memory available in the Apple computer, provision is made for handling as many as five variables with a total of 36 cells in the data matrix. The program is contained on a single 5-1/4 floppy disk, and may be run on an Apple II, Apple II+, or Apple IIe equipped with 64K or more of memory.

The programs on the disk provide for inputting the data from the keyboard, for computing the partition matrices defining an hierarchical model, for computing the expected values of the cells and appropriate statistics, and for testing the significance of the statistics. Elementary instructions for use of the program are included on the disk. A fuller description of the program is given in this report.

The purpose of this development has been to provide a capability for postulating and testing multivariate models of highway-safety-related data without the necessity of going to a large mainframe computer. The interactive nature of the microcomputer permits the sequential designation and testing of a variety of models in a period of only an hour or so.

17. Key Words	18. Distribution Statement

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
		22	

APPLE-VECTA' DOCUMENTATION

Phyllis A. Gimotty

September 1982

Introduction

Frequency tables are frequently used to summarize information about highway accidents. These tables are particularly appropriate when the variables of interest are qualitative classifications or ordered discrete categories but are also useful when continuous variables are "categorized" by assigning individuals to groups defined in terms of intervals of the continuous variable. While frequency tables present summarized data in a convenient way, it is often important to fit a model to the frequencies in the table to better understand the relationships among the variables defining the cells of the table.

This program is a convenient tool for analysis of frequency tables when there are more than two variables used to define the cells. The analysis is based on fitting a log-linear model to the cell frequencies. Hierarchical, as well as non-hierarchical, can be specified. The analysis provides information to describe the relationships among the variables that define the table by

---

'The APPLE-VECTA algorithm is based on VECTA, an algorithm, written in FORTRAN, developed by Dr. Keith Smith at the University of Michigan. APPLE-VECTA is written in PASCAL and is compatible with the APPLE PASCAL 1.1 compiler.

ordering and testing the relative importance of the interactions among variables.

Data Definition

The data are input into the program in the form of a frequency table[2]. The data need to be arranged so that the dependent variable defines the rows and all combinations of the relevant independent variables define the columns. The partition matrices are defined using only one variable that defines the rows.

In APPLE-VECTA, the frequency table is limited to 36 cells. This limit is not excessively restrictive. It will accommodate, for example, a frequency table defined by five variables each with two levels (32 cells). With three variables, the program can accommodate a table defined by two variables with four levels each and one variable with two levels (32 cells), or a table defined by two variables with three levels each and one with four levels (36 cells).

Models involving more than five variables are exceedingly complicated to work with and difficult to interpret. The number of cells in the table increases with the number of variables and the number of levels for each variable. As the number of cells gets larger, the table entries can become more sparse and there may be large numbers of cells where data were not observed. The statistical tests rely on large sample properties and therefore

---

[2]This feature makes APPLE-VECTA easy to use with published data that has already been tabulated. When case data are available, other analysis packages, such as OSIRIS, SAS, and SPSS, can be used to provide the tabulated data in a form appropriate for APPLE-VECTA.

2

may not be adequate with tables that have many cells with frequencies that are small.

The frequencies in each cell must be integers. The program will not accommodate frequencies from a weighted analysis resulting in non-integer frequencies. In this situation, the frequencies should be rounded to integers before entering the data. The statistical tests are appropriate for multinomial data, including data obtained from an equal probability sample. The G^2 statistic may not be appropriate for weighted data.

Table 1 is an example of data to be analyzed. When APPLE-VECTA is called from the main menu a brief introduction is printed, and then a transfer is made to a separate routine for data input. A <return> entry at the first point will start the transfer. You will be asked whether you wish to save an existing data set (the last one used), and a "y" response to this will lead to a prompt for entering the desired model definition. If new data are to be entered (an "n" response), you will be asked to enter (1) the number of variables (1 to 5), and then the number of levels (in turn) for each of the variables. These should be in the order of the appearance of variables on the table, beginning with the "A" or row variable, and following with "B", "C", etc. In Table 1 these entries would be "3" (for the number of variables) followed by 2, 3, and 4 (for the number of levels for the three variables shown).

The program will then ask for the data. The cell frequencies are entered by row; that is, the order for the data is: 37, 56, 73, 21, 83, 56, 72, 49, 101, 97, 35, 86. Input one

3

entry per line (return after typing each frequency--use no commas) then enter the second row in the same manner.

Table 1

An Example of a Frequency Table

	B=1				B=2				B=3			
	C=1	C=2	C=3	C=4	C=1	C=2	C=3	C=4	C=1	C=2	C=3	C=4
A=1	37	56	73	21	83	56	72	49	101	97	35	86
A=2	12	23	17	21	55	43	36	25	14	8	11	9

There is no provision for editing the data after they have been entered. Since the maximum number of cells which can be entered in the present APPLE version is 36, this is not viewed as a great hardship. If you make an error in entering the data, just restart the program and try again. After the data have been entered, the next step will be to define the log-linear model.

Model Definition

There are many references on the analysis of a frequency table using a log-linear model. Bishop, Feinberg, and Holland[3] give a fairly complete presentation of modeling a frequency table with a log-linear model. For this discussion, consider the case of a three-dimensional table defined by three variables, A, B, and C. The saturated log-linear model can be expressed as

$$\ln \pi_{ijk} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB} + \lambda_{ik}^{AC} + \lambda_{jk}^{BC} + \lambda_{ijk}^{ABC}$$

[3]Bishop, Y.M.M., Feinberg, S.F., and Holland P.W. [1975]. Discrete Multivariate Analysis. Massachusetts, The MIT Press.

4

where ln π_{ijk} is the natural logarithm of the true cell probability, π_{ijk}; λ is a constant; λ_i^A, λ_j^B, and λ_k^C are main effects of the three factors; λ_{ij}^{AB}, λ_{ik}^{AC}, and λ_{jk}^{BC} are the two-factor effects; and λ_{ijk}^{ABC} is the three-factor effect. The highest-order effect in the saturated model is the three-factor effect, λ_{ijk}^{ABC}.

For hierarchical models the partition matrices are generated automatically following your specification of the desired model with letter codes. For example, the model for all two-way combinations of the three variables (A, B, and C) would be specified by entering the string AB,AC,BC at the prompt for model definition. Note that a comma must be placed between the terms, but that there should be no comma at the end. The partition matrices used in the subsequent computation are generated from an analysis of this string.

In the following section of this documentation there is a discussion of the method of defining the partition matrices. It is not necessary to understand this method to run the program, and you may wish to skip the detailed discussion. However, a reading of that section should better enable you to understand the APPLE-VECTA algorithm.

There are also instructions available on the disk which will assist in understanding the procedure for inputting the model specification.

Partition Matrices

The model is specified by partition matrices. In the case of a three-dimensional frequency table, eight partition matrices are needed to specify the saturated model, one representing each term of Equation 1. A partition matrix consists of integers assigned to the cells of the data matrix.

The pattern of integers in the data matrix is defined by the main effect or multi-factor effect under consideration. The partition matrix for the three-factor effect (ABC) is defined in the following manner. First, it is necessary to find how many combinations of different levels of the three variables, A, B, and C, there are. This is determined by multiplying the number of levels of variable A times the levels of variable B times the levels of variable C. For our example, variable A has 2 levels, variable B has 3 levels, and variable C has 4 levels; therefore there are 24 different combinations of different values of the variables, A, B, and C. An integer is assigned to each possible combination of levels. The partition matrix for the ABC-effect has 24 distinct integers assigned to cells. Each cell is coded with the number associated with a unique combination of the levels of variables A, B, and C. The partition matrix, for the three-factor effect ABC, is given in Table 2.

The partition matrix for the two-factor effect involving variables A and B is given by the matrix in Table 3. Note that the number of levels of variable A is 2 and the number of levels of variable B is 3 and, therefore, there are six (2 x 3 = 6) combinations of levels of variables A and B. Each cell is identified with a unique integer corresponding to its level of

Table 2

Partition Matrix
for the ABC-Effect

	B=1				B=2				B=3			
	C=1	C=2	C=3	C=4	C=1	C=2	C=3	C=4	C=1	C=2	C=3	C=4
A=1	1	2	3	4	5	6	7	8	9	10	11	12
A=2	13	14	15	16	17	18	19	20	21	22	23	24

Table 3

Partition Matrix
for the AB-Effect

	B=1				B=2				B=3			
	C=1	C=2	C=3	C=4	C=1	C=2	C=3	C=4	C=1	C=2	C=3	C=4
A=1	1	1	1	1	2	2	2	2	3	3	3	3
A=2	4	4	4	4	5	5	5	5	6	6	6	6

the variables A and B. (For example, the integer 4 is assigned to those cells for which the level of variable A is 2 and the level of variable B is 1.). The program does not care which order the integers are assigned to the 6 combinations of the levels of the variables A and B. Either the partition matrix in Table 3 or the partition matrix in Table 4 could be used to represent the AB-effect.

Partition matrices for the remaining terms in the log-linear model are presented in Tables 5-9. A partition matrix for the

7

Table 4

Alternative Partition Matrix
for the AB-Effect

	B=1				B=2				B=3			
	C=1	C=2	C=3	C=4	C=1	C=2	C=3	C=4	C=1	C=2	C=3	C=4
A=1	1	1	1	1	3	3	3	3	5	5	5	5
A=2	2	2	2	2	4	4	4	4	6	6	6	6

two-factor effect defined by variables A and C (involving $2 \times 4 = 8$ integers) is given in Table 5. A partition matrix for the BC-effect (involving $3 \times 4 = 12$ integers) is given by Table 6. A partition matrix for the main effect of variable A (involving 2 integers) is given in Table 7, a partition matrix for the B main effect (involving 3 integers) is given by Table 8, and a partition matrix for the main effect of variable C (involving 4 integers) in given in Table 9. The partition matrix for the constant term in the model, λ, is a matrix with all elements equal to one. The program automatically creates a partition matrix for this term.

You need only specify the highest-order partition matrix (or matrices) to define the model. If the model to be fit is the saturated model, then the partition matrix associated with the three-factor effect is the only partition matrix that needs to be specified.[4] If the model consists of all three two-factor

---

[4]In practice the saturated model is not of interest since the predicted values of the cell probabilities are the observed

8

Table 5

Partition Matrix
for the AC-Effect

	B=1				B=2				B=3			
	C=1	C=2	C=3	C=4	C=1	C=2	C=3	C=4	C=1	C=2	C=3	C=4
A=1	1	2	3	4	1	2	3	4	1	2	3	4
A=2	5	6	7	8	5	6	7	8	5	6	7	8

Table 6

Partition Matrix
for the BC-Effect

	B=1				B=2				B=3			
	C=1	C=2	C=3	C=4	C=1	C=2	C=3	C=4	C=1	C=2	C=3	C=4
A=1	1	2	3	4	5	6	7	8	9	10	11	12
A=2	1	2	3	4	5	6	7	8	9	10	11	12

effects, the three matrices associated with the AB-, AC-, and AB-effect must be specified but <u>not</u> the main effects. If the model includes the AB- and BC-effect, then only those two partition matrices are needed. But, if the only effect specified by the model is the AB-effect, then it is necessary to use a partition matrix for the main effect of variable C.

proportions for each cell. Consequently, the saturated model will always fit the data exactly.

Table 7

Partition Matrix
for the A-Effect

	B=1				B=2				B=3			
	C=1	C=2	C=3	C=4	C=1	C=2	C=3	C=4	C=1	C=2	C=3	C=4
A=1	1	1	1	1	1	1	1	1	1	1	1	1
A=2	2	2	2	2	2	2	2	2	2	2	2	2

Table 8

Partition Matrix
for the B-Effect

	B=1				B=2				B=3			
	C=1	C=2	C=3	C=4	C=1	C=2	C=3	C=4	C=1	C=2	C=3	C=4
A=1	1	1	1	1	2	2	2	2	3	3	3	3
A=2	1	1	1	1	2	2	2	2	3	3	3	3

Table 9

Partition Matrix
for the C-Effect

	B=1				B=2				B=3			
	C=1	C=2	C=3	C=4	C=1	C=2	C=3	C=4	C=1	C=2	C=3	C=4
A=1	1	2	3	4	1	2	3	4	1	2	3	4
A=2	1	2	3	4	1	2	3	4	1	2	3	4

10

To summarize, the model must be specified in terms of main effects and multi-factor effects through partition matrices. The description of a model involves specifying all main effects and multi-factor effects to be included in the model. The number of partition matrices needed to describe the model are generally fewer than the number of main effects and multi-factor effects specified. To determine the partition matrices that are necessary, identify the highest-order effect (including the main effect, if necessary) for each variable and input those partition matrices.

Analysis Results

The algorithm used by APPLE-VECTA is an algorithm that combines iterative proportional fitting (IPF) and a Newton-Raphson search algorithm to find maximum likelihood estimates for the parameters in the log-linear model under the assumption of multinomial sampling. A discussion of this approach to fitting log-linear models can be found in Haberman[5].

Returning to the three-dimensional frequency table, if the model specifies that

$$\ln \pi_{ijk} = \lambda + \lambda_k^C + \lambda_{ij}^{AB} \tag{2}$$

then the algorithm uses the observed proportions to obtain estimates for the parameters $\hat{\lambda}$, $\hat{\lambda}_k^C$, and $\hat{\lambda}_{ij}^{AB}$. The estimated cell probabilities, $\hat{p}_{ijk}$, for the model specified by Equation 2 are then obtained by

$$\hat{p}_{ijk} = \exp(\hat{\lambda} + \hat{\lambda}_k^C + \hat{\lambda}_{ij}^{AB}). \tag{3}$$

[5]Haberman, S.J. [1974]. The Analysis of Frequency Data. Chicago, The University of Chicago Press.

where "exp" is the natural exponential function.

Two methods are available to examine how well the hypothesized model describes the observed proportions, p_{ijk}. A test of statistical significance of the model, G^2, is presented. G^2 is the likelihood ratio test statistic. This statistic has an asymptotic chi-square distribution and the degrees of freedom are calculated and printed along with the value of G^2.

A program is provided on the diskette to compute the statistical significance of the hypothesized model. To enter this program you should know both the value of the G^2 from the run and the number of degrees of freedom. Choosing (2) from the main menu will invoke a program which takes these two values as arguments and returns the significance level of that combination. While the interpretation of the significance level must be made with consideration of the sample size, very small values of this statistic indicate that the model does not fit the data well. For example, if the significance is less than, say, 0.001, the model should be rejected as a poor fit. On the other hand, if the significance is, say, 0.3, then the model fits the data rather well and this would become a candidate model for consideration. Several different models may fit the data, and one should seek the most parsimonious model (the one with the fewest factors and interactions) which explains the data reasonably well.

The G^2 statistic is used to see if the model is statistically significant. It is possible to have a model that is statistically significant but, yet, is not practically

significant. The Freeman-Tukey deviates, for each cell, are also printed as part of the output. These statistics, z_{ijk}, are a function of the observed and predicted probabilities and are defined by

$$z_{ijk} = (np_{ijk})^{\frac{1}{2}} + (np_{ijk} + 1)^{\frac{1}{2}} - (4n\hat{p}_{ijk} + 1)^{\frac{1}{2}}$$

where p_{ijk} is the observed proportion for the ijk^{th} cell, and $\hat{p}_{ijk}$ is the estimated cell probability for the ijk^{th} cell. If any of the deviates are greater than 1 in absolute value, then the model is doing a poor job in predicting the cell probability. Generally a model which produces a good fit (as measured by the G^2) will have very few cells (in the Freeman-Tukey deviate table) which are greater than 1 in absolute value. For a discussion of the Freeman-Tukey deviates consult Bishop, Feinberg, and Holland.[*]

The computations the program does take some time, especially if it is a complex model. After the model has been defined, there is no need for further input. Just wait for the output! Several messages will be displayed while the parameters are being estimated, and a dot is displayed for each iteration of several subroutines. Specifically, messages will include:

1. SETTING UP THE MODEL

2. FITTING THE MODEL

3. THE MODEL HAS BEEN FIT

4. THE RESULTS FOR [MODEL DEFINITION] ARE

---

[*]Bishop, Y.M.M., Feinberg, S.F., and Holland P.W. [1975]. Discrete Multivariate Analysis. Massachusetts, The MIT Press. See page 130.

Following the last message, the model definition, the degrees of freedom, G^2, the estimated cell frequencies based on the model, the original frequencies, and the Freeman-Tukey deviates are produced. Pressing <return> returns operation to the main menu from which you may rerun (with a different model), compute the significance level of the observed G^2, reread the instructions, or quit.

If you have a printer connected to the APPLE you may toggle the printer off or on from the main menu by pressing (5). When the printer has been invoked the model definition and the output tables will be printed there, and the tables will also appear on the screen. Although the printer command is normally off (when the disk is booted), toggling the printer on once should keep it on for a full session.

An Example

A three-dimensional table has been selected to demonstrate the use of APPLE-VECTA. Previous research[7] has suggested that injury severity and Delta V, a common measure of crash severity, are related. The physical model that the calculation of Delta V is based on involves a relationship between passenger car size and Delta V. Log-linear models can be used to see how these three variables interact with each other.

For convenience, the three variables will be referred to as A, B, and C, where A represents injury severity, B represents

---

[7]Gimotty, P.A., Campbell, K.L., and Chirachavala, T. [1980]. Statistical Analysis of the National Crash Severity Study. Washington, D.C., National Highway Traffic Safety Administration.

vehicle size, and C represents Delta V. Injury severity has been recoded into two categories: Severe (reported or estimated Overall Injury Severity, OAIS, of 4 to 6), and Non-severe (reported or estimated OAIS of 0 to 3). Passenger car size has been recoded into three vehicle size categories: Small (including compacts and smaller), Medium (including intermediate), and Large. Delta V has been recoded into three categories: L(ow), which is a Delta V less than or equal to 20 mph, M(edium), a Delta V between 20 mph and 40 mph, and H(igh), a Delta V greater than 40 mph.

The data used are from the NCSS Occupant File[8]. Each count in the frequency table represents one occupant. All of the occupants included in the table were in passenger cars involved in front collisions. This table includes only those occupants in the file for whom information was recorded for all three variables. The data are presented in Table 10. The log-linear model for these variables is given by

$$\ln \pi_{ijk} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB} + \lambda_{ik}^{AC} + \lambda_{jk}^{BC}$$

where π_{ijk} are the true cell probabilities; λ is a constant; λ_i^A, λ_j^B, and λ_k^C are main effects of the three factors; λ_{ij}^{AB}, λ_{ik}^{AC}, and λ_{jk}^{BC} are the two-factor effects; and λ_{ijk}^{ABC} is the three-factor effect.

---

[8] The NCSS (National Crash Severity Study) is a National Highway Traffic Safety Administration sponsored data base collected between January, 1977 and March, 1979, to study the relationship between injury severity and crash severity. The data used in this example was obtained from the Highway Safety Research Institute's version of that data set.

For this example, all possible models were fit and are described in Table 12. The model is described by the parameters included in the model. The model parameters, given in column one of Table 12, specify the model to be fit. The partition matrices needed to fit the model using APPLE-VECTA are given in column two. Notice that there are always fewer partition matrices than parameters in the model. For a further explanation see the discussion on Model Definition. The partition matrices are described, for this problem, in Table 11. The appropriate matrices from Table 11 are input into APPLE-VECTA when partition matrices are required.

Also in Table 12 are the statistical results of the model fitting. There is an ordering to the value of G^2 that is related to how well the model fits the data. The very large values of G^2 indicate that the model does not fit the data very well. They are those models that are indicated as significant. With a large value of G^2, the model significantly differs from the data.

There are different ways to look at the data presented in Table 10. Each of the three variables could, in different situations, be the main variable of interest. In regression terminology, this variable would be the "dependent variable." The remaining variables would be considered the "independent variables." The associations between the dependent variable and each independent variable are the most interesting associations since they describe the relationships between the dependent variable and each independent variable. Associations between independent variables are only of secondary interest. The

Table 10
Frequency Table of Injury Severity,
Car Size, and Delta V

INJURY SEVERITY (A)	VEHICLE SIZE (B)								
	SMALL VEHICLE			MEDIUM VEHICLE			LARGE VEHICLE		
	DELTA V (C)			DELTA V (C)			DELTA V (C)		
	L	M	H	L	M	H	L	M	H
Severe (OAIS 4-6)	8	43	22	33	72	40	42	84	47
Non-severe (OAIS 0-3)	601	174	19	1740	441	21	2342	540	35

results in Table 12 can be viewed from three different perspectives.

Firstly, if the objective of the research task is to examine injury severity, the dependent variable, as it relates to Delta V and vehicle size, the independent variables, we would interpret the results in Table 12 in the following manner. The first model in Table 12, which includes all the two-factor effects, is not significant (G^2 = 6.24, df = 4). That is, this model is consistent with the patterns in the observed data. We conclude from this result that any association between injury severity and Delta V is the same for occupants in small cars, occupants in medium cars, and occupants in large cars. The second model in Table 12, excluding the two-factor effect of injury severity and Delta V, was found to be significant (G^2 = 714.35, df = 6). We conclude that the two-factor effect of injury severity and

17

Table 11

Partition Matrices for Example

FACTOR	PARTITION MATRIX								
A	1	1	1	1	1	1	1	1	1
	2	2	2	2	2	2	2	2	2
B	1	1	1	2	2	2	3	3	3
	1	1	1	2	2	2	3	3	3
C	1	2	3	1	2	3	1	2	3
	1	2	3	1	2	3	1	2	3
AB	1	1	1	2	2	2	3	3	3
	4	4	4	5	5	5	6	6	6
AC	1	2	3	1	2	3	1	2	3
	4	5	6	4	5	6	4	5	6
BC	1	2	3	4	5	6	7	8	9
	1	2	3	4	5	6	7	8	9

Delta V is large and injury severity and Delta V are significantly associated. Finally, the third model, excluding the two-factor effect of injury severity and vehicle size, was found not significant (G^2 = 7.67, df = 6). We conclude that the two-factor effect of injury severity and vehicle size is zero and there is no association between injury severity and vehicle size.

Secondly, if the objective is to study vehicle size, the dependent variable, as it relates to injury severity and Delta V, the independent variables, we would look at the results in the following manner. From the first model, including all two-factor

Table 12

Summary of Analysis Results

MODEL PARAMETERS†	PARTITION MATRICES	DEGREES OF FREEDOM	G^2	SIGNIFICANCE LEVEL
Three-Variable Models				
$\lambda, \lambda_i^A, \lambda_j^B, \lambda_k^C, \lambda_{ij}^{AB}, \lambda_{ik}^{AC}, \lambda_{jk}^{BC}$	AB,AC,BC	4	6.24	0.1819
$\lambda, \lambda_i^A, \lambda_j^B, \lambda_k^C, \lambda_{ij}^{AB}, \lambda_{jk}^{BC}$	AB,BC	6	714.35	0.0000
$\lambda, \lambda_i^A, \lambda_j^B, \lambda_k^C, \lambda_{ik}^{AC}, \lambda_{jk}^{BC}$	AC,BC	6	7.67	0.2633
$\lambda, \lambda_i^A, \lambda_j^B, \lambda_k^C, \lambda_{ij}^{AB}, \lambda_{ik}^{AC}$	AB,AC	8	20.59	0.0083
$\lambda, \lambda_i^A, \lambda_j^B, \lambda_k^C, \lambda_{ij}^{AB}$	AB,C	10	735.92	0.0000
$\lambda, \lambda_i^A, \lambda_j^B, \lambda_k^C, \lambda_{ik}^{AC}$	AC,B	10	29.23	0.0011
$\lambda, \lambda_i^A, \lambda_j^B, \lambda_k^C, \lambda_{jk}^{BC}$	A,BC	8	722.90	0.0000
$\lambda, \lambda_i^A, \lambda_j^B, \lambda_k^C$	A,B,C	12	778.50	0.0000
Two-Variable Models				
$\lambda, \lambda_j^B, \lambda_k^C, \lambda_{jk}^{BC}$	BC	9	6530.82	0.0000
$\lambda, \lambda_i^A, \lambda_k^C, \lambda_{ik}^{AC}$	AC	12	1396.18	0.0000
$\lambda, \lambda_i^A, \lambda_j^B, \lambda_{ij}^{AB}$	AB	12	6455.59	0.0000
$\lambda, \lambda_i^A, \lambda_j^B$	A,B	14	8559.70	0.0000
$\lambda, \lambda_i^A, \lambda_k^C$	A,C	14	2111.46	0.0000

Table 12 (Continued)

Summary of Analysis Results

MODEL PARAMETERS†	PARTITION MATRICES	DEGREES OF FREEDOM	G^2	SIGNIFICANCE LEVEL
$\lambda, \lambda_j^B, \lambda_k^C$	B,C	13	6552.38	0.0000
One-Variable Models				
λ, λ_i^A	A	16	7831.13	0.0000
λ, λ_j^B	B	15	1227.21	0.0000
λ, λ_k^C	C	15	7919.28	0.0000

†A is Injury Severity.
B is Vehicle Size.
C is Delta V.

effects, we can conclude that any association between vehicle size and Delta V is the same for occupants severely injured and occupants non-severely injured. The third model, excluding the two-factor effect of vehicle size and injury severity is not significant (G^2 = 7.67, df = 6). We conclude the effect of the two-factor effect of vehicle size and injury severity is negligible and there is no association between vehicle size and injury severity. The fourth model, excluding the two-factor effect of vehicle size and Delta V, is significant (G^2 = 20.59, df = 8). We conclude that the effect of vehicle size and Delta V is large and there is an association between Delta V and vehicle size.

Finally, if the objective is to examine Delta V, the dependent variable, as it relates to injury severity and vehicle size, the independent variables, we would interpret the results in Table 12 in the following manner. The first model in Table 12, including all two-factor effects, is not significant (G^2 = 6.24, df = 4). We conclude from this result that any association between Delta V and vehicle size is the same for occupants severely injured as for occupants not severely injured. The second model in Table 12, excluding the two-factor effect of injury severity and Delta V, is significant (G^2 = 714.35, df = 6). We conclude that the two-factor effect of injury severity and Delta V is large and Delta V and injury severity are significantly associated. Finally, the fourth model, excluding the two-factor effect of Delta V and vehicle size, was found to be significant (G^2 = 20.59, df = 8). We conclude that the two-

factor effect of Delta V and vehicle size is large and Delta V and injury severity are also significantly associated.

A general statement can be made about the relationships in Table 12. The cell probabilities, for this table, are best explained by the association between injury severity and Delta V and the association between Delta V and vehicle size. The two-factor effect representing an association between injury severity and car size does not appear in the best fitting model because of the lack of association between injury severity and vehicle size.

Sample Data

A set of data to demonstrate interaction among four variables is described in the instructions on the APPLE.VECTA diskette, and is recorded in a file named "SAMPLE.DATA."

The data file used by the program is always named VEC1, and you may transfer the sample file to VEC1 by quitting the menu (press "6"), calling the Pascal Files (press "F"), and entering the sequence:

```
T                  [for Transfer]
SAMPLE.DATA        [file to be transferred]
VEC1               [Destination]
```

Pressing "Q" after this has been accomplished will return to the Pascal operating system. You may restart by pressing "H" to reboot or by executing a program named START.