

A COMPUTER PROGRAM FOR TESTING AVERAGE PARTIAL ASSOCIATION IN THREE-WAY CONTINGENCY TABLES (PARCAT)

J. Richard LANDIS, Murray M. COOPER, Thomas KENNEDY⁺ and Gary G. KOCH⁺

Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109 and ⁺ Department of Biostatistics, University of North Carolina, Chapel Hill, NC 27514, USA

PARCAT is a computer program which implements alternative tests for average partial association in three-way contingency tables within the framework of the product multiple hypergeometric probability model. Primary attention is directed at the relationship between two of the variables, controlling for the effects of a covariable. This approach is essentially a multivariate extension of the Cochran/Mantel-Haenszel test to sets of $(s \times r)$ tables. A set of scores such as uniform, ridsits, or probits can be assigned to categories which are ordinally scaled. In particular, if ridit scores with midranks assigned for ties are utilized, this procedure is equivalent to a partial Kruskal-Wallis test when one variable is ordinally scaled, and is equivalent to a partial Spearman rank correlation test when both variables are ordinally scaled.

Categorical data Partial association Cochran/Mantel-Haenszel tests Hypergeometric probability model

1. Introduction

For research situations in which the basic data are measured in terms of discrete categories (based on nominal or ordinal as opposed to interval scales) the primary question of interest frequently involves the relationship between a set of s categories which correspond to distinct sub-populations (as defined in terms of pertinent independent variables) and a set of r categories which correspond to the response profiles associated with the specific dependent variables under study. In addition, the distribution of the response profiles may also be influenced by the effects of a secondary set of q categories which correspond to distinct levels of relevant covariables such as investigators, hospitals, clinics, or pretreatment states. As a result, the data obtained from such studies can be summarized in a set of q : $(s \times r)$ contingency tables which will be indexed by $h = 1, 2, \dots, q$. In this formulation, the basic hypothesis can then be expressed in terms of 'no partial association' between the sub-populations and the response profiles, after adjusting for the possible effects of the covariables. For example, suppose either treatment A or B is assigned at random to subjects within a series of q clinics and the response to treatment is classified as either improved

or not improved. In this situation, the resulting data can be displayed in a set of q : (2×2) tables for which the primary question is whether, on the average, across the clinic subgroups, the improvement rates are the same for the two treatments or are consistently different.

Cochran [1] proposed a test statistic for this hypothesis with respect to a set of (2×2) tables from the point of view of asymptotic binomial model results (which require moderately large sample sizes, e.g., $n_h \geq 20$ for $h = 1, 2, \dots, q$). Alternatively, Mantel and Haenszel [2] noted that this same problem could be approached within the framework of a hypergeometric model which requires only the overall sample size $n = \sum_h n_h$ to be large for asymptotic methods to be applicable. In fact, their procedure is appropriate for matched case-control studies with only two subjects in each of the q tables. More specifically, this method utilizes expected values and variances for a specified cell (pivot cell) in each of the (2×2) tables but is invariant among which of the four cells is selected. However, since the structure of their statistic differs from Cochran's statistic only by a factor of $[(n_h - 1)/n_h]$ in the variance term for the pivot cell of each of the q tables and by a continuity correction, the results obtained from these two methods are

essentially equivalent when the sample sizes in each table are moderate to large (e.g., $n_h \geq 20$ for $h = 1, 2, \dots, q$). For other cases, the Mantel-Haenszel method is preferred because it only requires asymptotic considerations on an across-table basis rather than both across and within tables.

More generally, Mantel and Haenszel [2] indicated that their procedure could be extended to investigate the concept of 'average partial association' for a set of $q: (2 \times r)$ contingency tables. This method involves computing the expected values and the covariance matrix for $(r - 1)$ pivot cells under the multiple hypergeometric probability model for each of the tables. These quantities are then summed across the q tables and a quadratic form test statistic is generated. This same procedure was also discussed by Birch [3] and more recently by Sugiura and Otake [4]. However, in Mantel and Haenszel [2] the specific computations for this method were outlined only for a set of (2×3) contingency tables. Rather than developing the details for an $(s - 1)(r - 1)$ degree of freedom test statistic in general, Mantel [5] presented such results for the situation in which the response profiles are ordinally scaled with progressively larger intensities. For this case, he discussed various scoring techniques and proposed a test statistic based on the resulting mean scores. Moreover, if both dimensions of the $(s \times r)$ contingency tables are ordinally scaled, Mantel [5] recommended score correlation-type test statistics. Otherwise, the multivariate Mantel-Haenszel procedure involving $(s - 1)(r - 1)$ pivot cells has been used in recent years in the analysis of highway safety data, as demonstrated in [6].

In a recent paper, Landis, Heyman and Koch [7] summarized various alternative approaches for investigating the underlying concept of 'average partial association' in three-way contingency tables. In particular, they presented a unified notation and matrix formulation for the Generalized Cochran/Mantel-Haenszel (CMH) approach to the analysis of $q: (s \times r)$ contingency tables in terms of the corresponding multiple hypergeometric probability model. The purpose of this paper is to describe a new computer program PARCAT which implements this generalized CMH hypothesis testing framework for the analysis of three-way contingency tables. The statistical methodology for this procedure is outlined in section 2 and the possible scoring options for ordinally scaled variables

are discussed in section 3. Specifically, if ridit scores with midranks assigned for ties are utilized, this procedure is equivalent to a partial Kruskal-Wallis test when the response variable is ordinally scaled and is equivalent to a partial Spearman rank correlation test when both dimensions of each $(s \times r)$ contingency table are ordinally scaled.

The specific details associated with data input are outlined in section 4 and the description of the control cards used in the execution of PARCAT is summarized in considerable detail in section 5. Finally, the analyses of two typical data sets using PARCAT are presented in section 6 with particular attention given to the required control cards.

Several analyses of a given data set can be performed in the same computer run by specifying additional sets of scores for either the row or column dimensions. Furthermore, multiple data sets can be processed in the same computer run by simply repeating the appropriate sequence of control cards, together with new data, as described further in section 5.

2. Methodology

Let $h = 1, 2, \dots, q$ index a set of $(s \times r)$ contingency tables which correspond to distinct levels of a covariable or combinations of several pertinent covariables. Let $i = 1, 2, \dots, s$ index a set of sub-populations which are to be compared with respect to a particular response variable for which the outcome categories are indexed by $j = 1, 2, \dots, r$. Then let $\mathbf{n}'_h = (n_{h11}, \dots, n_{h1r}, \dots, n_{hs1}, \dots, n_{hsr})$, where n_{hij} denotes the number of subjects in the sample who are jointly classified as belonging to the h -th table, the i -th sub-population and the j -th response category. These frequency data, corresponding to the h -th level of the covariable set, can be summarized as shown in table 1, where N_{hi} denotes the marginal total number of subjects classified as belonging to the i -th sub-population, $N_{h \cdot j}$ denotes the marginal total number of subjects classified as belonging to the j -th response category, and $N_{h \cdot \cdot}$ denotes the overall marginal total sample size in the h -th table.

The basic hypothesis under investigation involves the relationship between the response variable and the sub-populations adjusted for the levels of the

Table 1
Observed contingency table for level h of the covariables

Sub-population	Response variable categories				Total
	1	2	...	r	
1	n_{h11}	n_{h12}	...	n_{h1r}	$N_{h1\cdot}$
2	n_{h21}	n_{h22}	...	n_{h2r}	$N_{h2\cdot}$
⋮	⋮	⋮	⋮	⋮	⋮
s	n_{hs1}	n_{hs2}	...	n_{hsr}	$N_{hs\cdot}$
Total	$N_{h\cdot 1}$	$N_{h\cdot 2}$...	$N_{h\cdot r}$	$N_{h\cdot\cdot}$

covariable set. Under the assumption that the marginal totals $\{N_{hi\cdot}\}$ and $\{N_{h\cdot j}\}$ are fixed (either by design or conditional distribution arguments), the overall null hypothesis of ‘no partial association’ can be stated as

H_0 : For each of the separate levels of the covariable set $h = 1, 2, \dots, q$, the response variable is distributed at random with respect to the sub-populations, i.e., the data in the respective rows of the h -th table can be regarded as a successive set of simple random samples of sizes $\{N_{hi\cdot}\}$ from a fixed population corresponding to the marginal total distribution of the response variable $\{N_{h\cdot j}\}$ (2.1)

On the basis of this hypothesis, it can be shown from stratified sampling arguments that the vector \mathbf{n}_h follows the product multiple hypergeometric distribution given by the probability model:

$$Pr(\mathbf{n}_h | H_0) = \frac{\prod_{i=1}^s N_{hi\cdot}! \prod_{j=1}^r N_{h\cdot j}!}{N_{h\cdot\cdot}! \prod_{i=1}^s \prod_{j=1}^r n_{hij}!} \tag{2.2}$$

Given this background, the types of applications where the general model (2.2) is of interest include:

- a. Observational and/or historical data from restricted populations as obtained in, e.g., retrospective studies, prospective non-randomized studies and case-control studies. Here, of

course, all of the frequency counts are fixed from a strict point of view because of the historical nature of the data. On the other hand, one can argue that there is still interest in the hypothetical question of whether or not the observed distribution of response is at random with respect to the sub-populations which are to be compared. Thus, it is through the inherent nature of this hypothesis itself that the $\{n_{hij}\}$ are regarded as random variables for such situations.

- b. Experimental design data from a strict randomization model point of view. Here, the marginal distributions of sub-populations are regarded as fixed in principle by the nature of the experimental design (but may actually be subject to some inherent variability because of missing data). In addition, the marginal distribution of response is regarded as fixed for the situation where there is no relationship between sub-population and response (since the response of each subject is not affected by the randomly assigned sub-population) which implies that the distribution of their responses as a whole when combined across sub-populations, is the same for all realizations of the treatment randomization process.
- c. Product multinomial model data as discussed in Landis et al. [7] from either stratified simple random samples or stochastic processes with respect to a conditional distribution point of view.

For each table indexed by $h = 1, 2, \dots, q$, let $P_{hi\cdot} = (N_{hi\cdot}/N_{h\cdot\cdot})$ denote the marginal proportion of subjects classified as belonging to the i -th sub-population (row) and let $P_{h\cdot j} = (N_{h\cdot j}/N_{h\cdot\cdot})$ denote the marginal proportion of subjects classified as belonging to the j -th response category (column). These proportions, which are assumed to be fixed, can be summarized in vector notation for $h = 1, 2, \dots, q$ as $\mathbf{P}'_{h\cdot\cdot} = (P_{h1\cdot}, \dots, P_{hs\cdot})$ and $\mathbf{P}_{h\cdot\cdot} = (P_{h\cdot 1}, \dots, P_{h\cdot r})$. Then by computing the first and second moments of the probability distribution in (2.2), it can be shown that the expected value of n_{hij} under H_0 in (2.1) is $m_{hij} = N_{h\cdot\cdot} P_{hi\cdot} P_{h\cdot j}$. Thus, the expected value of \mathbf{n}_h can be expressed in vector notation as:

$$\mathbf{m}_h = E\{\mathbf{n}_h | H_0\} = N_{h\cdot\cdot} [\mathbf{P}'_{h\cdot\cdot} \otimes \mathbf{P}_{h\cdot\cdot}] \tag{2.3}$$

where \otimes denotes Kronecker product multiplication. Furthermore, the covariance between n_{hij} and $n_{hi'j'}$ under H_0 in (2.1) is:

$$V_{h,ij,i'j'} = N_{hi'}N_{h'j}(\delta_{ii'}N_{h..} - N_{hi'}) \times (\delta_{jj'}N_{h..} - N_{h'j'})/N_{h..}^2(N_{h..} - 1)$$

where $\delta_{ii'} = 1$, if $i = i'$, = 0, otherwise and $\delta_{jj'} = 1$, if $j = j'$, = 0, otherwise. Thus, the covariance matrix of \mathbf{n}_h under H_0 in (2.1) can be expressed in matrix notation as:

$$\mathbf{Var}\{\mathbf{n}_h|H_0\} = \frac{N_{h..}^2}{(N_{h..} - 1)} \{ [D_{P_{h^*}} - P_{h^*}P_{h^*}'] \otimes [D_{P_{h^*}} - P_{h^*}P_{h^*}'] \} \tag{2.4}$$

where $D_{P_{h^*}}$ and $D_{P_{h^*}}$ are diagonal matrices with elements of the vectors P_{h^*} and P_{h^*} on the main diagonal.

2.1. Multivariate test: case I

For the situation in which both dimensions of the $(s \times r)$ table represent data measured on nominal scales, the basic hypothesis in (2.1) can be tested in terms of $(s - 1)(r - 1)$ linearly independent functions of \mathbf{n}_h . Without loss of generality, let:

$$\mathbf{A} = \{ [I_{(r-1)}, \mathbf{0}_{(r-1)}] \otimes [I_{(s-1)}, \mathbf{0}_{(s-1)}] \} \tag{2.5}$$

where I_u denotes a $(u \times u)$ identity matrix and $\mathbf{0}_u$ denotes a $(u \times 1)$ vector of 0's. Thus for $h = 1, 2, \dots, q$, the quantities:

$$\mathbf{G}_h = \mathbf{A}(\mathbf{n}_h - \mathbf{m}_h) \tag{2.6}$$

represent the differences between the observed and expected frequencies under H_0 for the $(s - 1)(r - 1)$ cells in table 1, after eliminating the last row and column. From (2.3) and (2.4) it follows that:

$$\mathbf{E}\{\mathbf{G}_h|H_0\} = \mathbf{0}_{(s-1)(r-1)} \tag{2.7}$$

$$\mathbf{Var}\{\mathbf{G}_h|H_0\} = \frac{N_{h..}^2}{(N_{h..} - 1)} \{ [D_{\tilde{P}_{h^*}} - \tilde{P}_{h^*}\tilde{P}_{h^*}'] \otimes [D_{\tilde{P}_{h^*}} - \tilde{P}_{h^*}\tilde{P}_{h^*}'] \} \tag{2.8}$$

where $\tilde{P}_{h^*} = (P_{h1}, \dots, P_{h,(s-1)})$ and $\tilde{P}'_{h^*} = (P_{h1}, \dots, P_{h,(r-1)})$.

Given this framework in (2.6)–(2.8), it can be

shown that:

$$Q_h = \mathbf{G}'_h [\mathbf{Var}\{\mathbf{G}_h|H_0\}]^{-1} \mathbf{G}_h = \frac{(N_{h..} - 1)}{N_{h..}} \left\{ \sum_{i=1}^s \sum_{j=1}^r \frac{(n_{hij} - m_{hij})^2}{m_{hij}} \right\} = \frac{(N_{h..} - 1)}{N_{h..}} Q_{P,h} \text{ for } h = 1, 2, \dots, q \tag{2.9}$$

where $Q_{P,h}$ denotes the Pearson chi-square statistic for testing independence in the h -th table. Under H_0 in (2.1), Q_h asymptotically has the chi-square distribution with $d.f. = (s - 1)(r - 1)$. As a result:

$$Q_T = \sum_{h=1}^q \mathbf{G}'_h [\mathbf{Var}\{\mathbf{G}_h|H_0\}]^{-1} \mathbf{G}_h = \sum_{h=1}^q \sum_{i=1}^s \sum_{j=1}^r \left[\frac{N_{h..} - 1}{N_{h..}} \right] \left[\frac{(n_{hij} - m_{hij})^2}{m_{hij}} \right] \tag{2.10}$$

represents an appropriate test statistic for H_0 . In particular, if each of the respective $N_{h..}$ for $h = 1, 2, \dots, q$ is sufficiently large that all of the $\{Q_h\}$ have approximate chi-square distributions, then Q_T has an approximate chi-square distribution with $d.f. = q(s - 1) \times (r - 1)$ under (2.1).

On the other hand, if the overall sample size $N = \sum_h N_{h..}$ is large but the individual sample sizes $\{N_{h..}\}$ for many tables are small, then the statistic Q_T in (2.10) is no longer useful for testing H_0 in (2.1). In this case, it is more appropriate to investigate H_0 in terms of the sum of corresponding differences between the observed and expected frequencies across the q tables denoted by:

$$\mathbf{G} = \sum_{h=1}^q \mathbf{G}_h \tag{2.11}$$

From (2.7) and (2.8) it follows that:

$$\mathbf{E}\{\mathbf{G}|H_0\} = \mathbf{0}$$

$$\mathbf{Var}\{\mathbf{G}|H_0\} = \sum_{h=1}^q \mathbf{Var}\{\mathbf{G}_h|H_0\} \tag{2.12}$$

Consequently, a Generalized Cochran/Mantel-Haenszel statistic for testing H_0 can be obtained as:

$$Q_{CMH} = \mathbf{G}' [\mathbf{Var}\{\mathbf{G}|H_0\}]^{-1} \mathbf{G} \tag{2.13}$$

which asymptotically has the chi-square distribution with $d.f. = (s - 1)(r - 1)$. This test statistic is directed at 'average partial association' alternatives in the sense that if certain observed frequencies consistently exceed (or are exceeded by) their corresponding expected frequencies, across the tables indexed by $h = 1, 2, \dots, q$, then these quantities reinforce one another when combined to form Q_{CMH} . Also, the fact that the significance of Q_{CMH} is evaluated relative to $d.f. = (s - 1)(r - 1)$ rather than $q(s - 1)(r - 1)$ represents another aspect of this method that potentially permits gains in statistical power, relative to the overall partial association statistic Q_T in (2.10). Finally, the range of applications for Q_{CMH} is broader since its asymptotic chi-square distribution is linked to the overall sample size N , while that of Q_T is linked to the sample sizes $\{N_{h..}\}$ for each of the respective tables.

For the special case in which $s = r = 2$, the resulting data can be summarized in a set of q : (2×2) tables. In particular, G_h in (2.6) simplifies to:

$$G_h = \left(n_{h11} - \frac{N_{h1.} \cdot N_{h.1}}{N_{h..}} \right) \tag{2.14}$$

and the covariance matrix in (2.8) simplifies to the hypergeometric variance:

$$\begin{aligned} \text{Var}\{G_h|H_0\} &= \frac{N_{h..}^2}{(N_{h..} - 1)} P_{h1.}(1 - P_{h1.})P_{h.1}(1 - P_{h.1}) \\ &= \frac{N_{h1.} \cdot N_{h2.} \cdot N_{h.1} \cdot N_{h.2}}{N_{h..}^2 \cdot (N_{h..} - 1)} \end{aligned} \tag{2.15}$$

Thus, the test statistic for average partial association in (2.13) simplifies to:

$$Q_{CMH} = \frac{\left\{ \sum_{h=1}^q \left(n_{h11} - \frac{N_{h1.} \cdot N_{h.1}}{N_{h..}} \right) \right\}^2}{\left\{ \sum_{h=1}^q \frac{N_{h1.} \cdot N_{h2.} \cdot N_{h.1} \cdot N_{h.2}}{N_{h..}^2 \cdot (N_{h..} - 1)} \right\}} \tag{2.16}$$

This result in (2.16) differs from the test statistic proposed in Cochran [1] only by a factor of $[(N_{h..} - 1)/N_{h..}]$ in the variance term for each table and is identical (except for the lack of a continuity correction) to the statistic recommended in Mantel and Haenszel [2].

2.2. Mean score test: case II

For situations where the response categories $j = 1, 2, \dots, r$ are ordinally scaled with progressively larger intensities, it is often useful to test H_0 in terms of mean scores based on an appropriate vector of scores for the h -th table which can be denoted by $\mathbf{a}'_h = (a_{h1}, a_{h2}, \dots, a_{hr})$ for $h = 1, 2, \dots, q$. The particular choice of scores will not be discussed further here, but a detailed summary of various scoring procedures is outlined in section 3. In this context, let the $(s \times 1)$ vector of mean scores for the h -th table be denoted by $\mathbf{F}'_h = (F_{h1}, F_{h2}, \dots, F_{hs})$, where:

$$F_{hi} = \frac{1}{N_{h.}} \sum_{j=1}^r a_{hj} n_{hij} \tag{2.17}$$

is the mean score with respect to \mathbf{a}_h for the i -th subpopulation in the h -th table. Similarly, let the corresponding mean score on the column marginal proportions be denoted by:

$$\bar{a}_h = \sum_{j=1}^r a_{hj} P_{h.j} \tag{2.18}$$

Then from (2.3) it follows that:

$$\mathbf{E}\{\mathbf{F}_h|H_0\} = (\bar{a}_h \otimes \mathbf{1}_s) \tag{2.19}$$

where $\mathbf{1}_s$ is a vector of s 1's. Moreover, let the total variance of the response variable in the overall population for the h -th table with respect to \mathbf{a}_h be denoted by:

$$S_{a,h}^2 = \sum_{j=1}^r (a_{hj} - \bar{a}_h)^2 P_{h.j} \tag{2.20}$$

Then from (2.4) it follows that:

$$\text{Var}\{\mathbf{F}_h|H_0\} = \frac{S_{a,h}^2}{(N_{h..} - 1)} (\mathbf{D}_{\mathbf{F}_h}^{-1} - \mathbf{1}_s \mathbf{1}'_s) \tag{2.21}$$

Given this framework in (2.17)–(2.21), H_0 in (2.1) can be tested in terms of $(s - 1)$ contrasts among the s elements of \mathbf{F}_h . Without loss of generality, let $\mathbf{C} = [\mathbf{I}_{(s-1)}, -\mathbf{1}_{(s-1)}]$ be a basis for the space of contrasts among the estimators \mathbf{F}_h . Consequently, from (2.19) and (2.21) it follows that:

$$\mathbf{E}\{\mathbf{C}\mathbf{F}_h|H_0\} = \mathbf{0} \tag{2.22}$$

$$\text{Var}\{CF_h|H_0\} = \frac{S_{a,h}^2}{(N_{h\cdot\cdot} - 1)} \{CDP_{h^*}^{-1}C'\} \quad (2.23)$$

Thus, an appropriate test statistic for H_0 which is directed at alternatives pertaining to differences among the sub-populations with respect to the mean scores based on a_h is:

$$\begin{aligned} Q_{MS,h} &= (CF_h)' [\text{Var}\{CF_h|H_0\}]^{-1} CF_h \\ &= \frac{(N_{h\cdot\cdot} - 1)}{S_{a,h}^2} F_h' C' \{CDP_{h^*}^{-1}C'\}^{-1} CF_h \\ &= \frac{(N_{h\cdot\cdot} - 1)S_{F,h}^2}{S_{a,h}^2} \text{ for } h = 1, 2, \dots, q \end{aligned} \quad (2.24)$$

where

$$S_{F,h}^2 = \sum_{i=1}^s (F_{hi} - \bar{a}_h)^2 P_{hi}$$

Under H_0 in (2.1), $Q_{MS,h}$ asymptotically has the chi-square distribution with *d.f.* = ($s - 1$). As formulated in (2.24), $Q_{MS,h}$ can be viewed as essentially a one-way ANOVA test statistic for H_0 in the sense of being the ratio of the ‘Among sub-populations sums of squares’ versus the ‘Total sums of squares’. As a result, an overall test statistic for the simultaneous comparison of the mean scores within all q tables is:

$$Q_{MS} = \sum_{h=1}^q Q_{MS,h} \quad (2.25)$$

If each of the respective $N_{h\cdot\cdot}$ for $h = 1, 2, \dots, q$ is sufficiently large that all of the $\{Q_{MS,h}\}$ have approximate chi-squares distributions, then Q_{MS} has an approximate chi-square distribution with *d.f.* = $q(s - 1)$ under H_0 .

On the other hand, if the overall sample size N is large but the individual sample sizes $\{N_{h\cdot\cdot}\}$ for many tables are small, then the statistic Q_{MS} in (2.25) is no longer useful for testing H_0 in (2.1). In this case, it is more appropriate to investigate H_0 in terms of a Generalized Cochran/Mantel-Haenszel strategy with respect to the mean scores $\{F_h\}$ based on $\{a_h\}$. For this purpose, let the weighted sum of the mean scores across the q tables be denoted by:

$$F = \sum_{h=1}^q N_{h\cdot\cdot} DP_{h^*} F_h \quad (2.26)$$

Then it follows that:

$$E\{F|H_0\} = \sum_{h=1}^q N_{h\cdot\cdot} \bar{a}_h P_{h^*} \quad (2.27)$$

$$\text{Var}\{F|H_0\} = \sum_{h=1}^q \frac{N_{h\cdot\cdot}^2}{(N_{h\cdot\cdot} - 1)} S_{a,h}^2 \{DP_{h^*} - P_{h^*}P_{h^*}'\} \quad (2.28)$$

Due to the implicit matrix singularity in (2.28) which follows from the results that:

$$E\{\mathbf{1}'_s F|H_0\} = \sum_{h=1}^q N_{h\cdot\cdot} \bar{a}_h$$

$$\text{Var}\{\mathbf{1}'_s F|H_0\} = 0$$

the matrix C can be used as a basis for the space of contrasts among the estimators F . Consequently, by letting $D = C[F - E\{F|H_0\}]$ be the corresponding $[(s - 1) \times 1]$ vector of differences between the mean scores and their expected values under H_0 , the Generalized Cochran/Mantel-Haenszel statistic for H_0 in terms of mean scores with respect to $\{a_h\}$ is obtained as:

$$Q_{CMMS} = D' \{C[\text{Var}\{F|H_0\}]C'\}^{-1} D \quad (2.29)$$

Under H_0 , Q_{CMMS} asymptotically has the chi-square distribution with *d.f.* = ($s - 1$). This test statistic is directed at location shift alternatives which correspond to the extent to which the mean scores for certain sub-populations consistently exceed (or are exceeded by) the mean scores for other sub-populations; and thus it is directed at the average across the q tables of such differences. In particular, for $s = 2$, Q_{CMMS} in (2.29) is identical to the extended Mantel-Haenszel test statistic proposed in Mantel [5]. Furthermore, for the case where all the scores are either 0 or 1 and $s = 2$, Q_{CMMS} simplifies to the Cochran/Mantel-Haenszel statistic in (2.16) which corresponds to the collapsed set of (2×2) tables which are obtained by pooling the response categories which are assigned ‘0’ and those which are assigned ‘1’. Finally, it should be noted that if marginal rank or ridit-type scores are obtained from each table with midranks assigned for ties, the statistic Q_{CMMS} in (2.29) can be viewed as essentially equivalent to a partial Kruskal-Wallis ANOVA test on ranks, conditioning on the levels of the covariable set. Otherwise, the same types of

comments could be made about power properties and central limit theory considerations for these statistics as were made for the multivariate test in section 2.1, except that the concept of power is stronger relative to the location shift alternatives at which they are targeted and sample size requirements are weaker because fewer degrees of freedom are involved.

2.3. Correlation test: case III

For situations where both the response categories $j = 1, 2, \dots, r$ and the sub-population categories $i = 1, 2, \dots, s$ are ordinally scaled with progressively larger intensities, it is often useful to test H_0 in terms of composite mean scores which correspond to the respective products of an appropriate vector of response scores $\mathbf{a}'_h = (a_{h1}, a_{h2}, \dots, a_{hr})$ and an appropriate vector of sub-population scores $\mathbf{c}'_h = (c_{h1}, c_{h2}, \dots, c_{hs})$ for $h = 1, 2, \dots, q$. Specifically, let $w_{hij} = c_{hi}a_{hj}$ be the score assigned to the joint outcome corresponding to the i -th sub-population and j -th response category in the h -th table which can be summarized in vector notation by letting $\mathbf{w}'_h = (w_{h11}, \dots, w_{h1r}, \dots, w_{hs1}, \dots, w_{hsr})$. Then let the composite mean score function for the h -th table be defined by:

$$F_h = \mathbf{w}'_h \mathbf{n}_h / N_{h..} \tag{2.30}$$

Moreover, let the mean score on the row marginal proportions be denoted by:

$$\bar{c}_h = \sum_{i=1}^s c_{hi} P_{hi} \tag{2.31}$$

and let the mean score on the column marginal proportions be as defined in (2.18). Also, let the total variance of the sub-population variable in the overall population for the h -th table with respect to \mathbf{c}_h be denoted by:

$$S_{c,h}^2 = \sum_{i=1}^s (c_{hi} - \bar{c}_h)^2 P_{hi} \tag{2.32}$$

and let the corresponding total variance of the response variable be given as in (2.20). In addition, let:

$$S_{ac,h} = \sum_{i=1}^s \sum_{j=1}^r (a_{hj} - \bar{a}_h)(c_{hi} - \bar{c}_h)(n_{hij}/N_{h..}) \tag{2.33}$$

be the covariance between the response scores and the sub-population scores in the h -th table. Then from (2.3) and (2.4) it follows that:

$$E\{F_h | H_0\} = \bar{a}_h \bar{c}_h \tag{2.34}$$

$$\text{Var}\{F_h | H_0\} = S_{a,h}^2 S_{c,h}^2 / (N_{h..} - 1) \tag{2.35}$$

Furthermore, from (2.30) and (2.33) it can be shown that:

$$F_h - E\{F_h | H_0\} = S_{ac,h} \tag{2.36}$$

Given this framework in (2.30)–(2.36), H_0 in (2.1) can be tested in terms of a test statistic which is directed at alternatives pertaining to the association between the response variable and the sub-population variable under the composite scores $\{w_{hij}\}$. In particular, let:

$$\begin{aligned} Q_{MA,h} &= \frac{[F_h - E\{F_h | H_0\}]^2}{\text{Var}\{F_h | H_0\}} \\ &= \frac{(N_{h..} - 1) S_{ac,h}^2}{S_{a,h}^2 S_{c,h}^2} \\ &= (N_{h..} - 1) R_{ac,h}^2 \end{aligned} \tag{2.37}$$

where $R_{ac,h}^2$ denotes the squared Pearson correlation coefficient corresponding to the observed bivariate distribution for the $\{a_{hj}\}$ and the $\{c_{hi}\}$ in the h -th table. Under H_0 in (2.1), $Q_{MA,h}$ asymptotically has the chi-square distribution with $d.f. = 1$. As defined in (2.37), $Q_{MA,h}$ is essentially a standard correlation analysis test statistic for H_0 . As a result, an overall test statistic for the simultaneous test of the measures of association $\{F_h\}$ in (2.30) across all q tables is:

$$Q_{MA} = \sum_{h=1}^q Q_{MA,h} \tag{2.38}$$

If each of the respective $N_{h..}$ for $h = 1, 2, \dots, q$ is sufficiently large that all of the $\{Q_{MA,h}\}$ have approximate chi-square distributions, the Q_{MA} has an approximate chi-square distribution with $d.f. = q$ under H_0 in (2.1).

Alternatively, if the overall sample size N is large but the individual sample sizes $\{N_{h..}\}$ for many tables are small, then the statistic Q_{MA} in (2.38) is no longer useful for testing H_0 in (2.1). Instead it is more appropriate to investigate H_0 in terms of a Generalized Cochran/Mantel-Haenszel strategy with respect

to the measures of association $\{F_h\}$ in (2.30) based on the composite mean scores $\{\mathbf{w}_h\}$. For this purpose, let the weighted sum of the composite mean scores across the q tables be denoted by:

$$F = \sum_{h=1}^q N_{h\cdot\cdot} F_h \quad (2.39)$$

From (2.34) and (2.35) it follows that:

$$E\{F|H_0\} = \sum_{h=1}^q N_{h\cdot\cdot} \bar{a}_h \bar{c}_h \quad (2.40)$$

$$\text{Var}\{F|H_0\} = \sum_{h=1}^q \frac{N_{h\cdot\cdot}^2}{(N_{h\cdot\cdot} - 1)} S_{a,h}^2 S_{c,h}^2 \quad (2.41)$$

Consequently, the Generalized Cochran/Mantel-Haenszel statistic for H_0 in terms of composite mean scores with respect to \mathbf{w}_h is obtained as:

$$Q_{\text{CMMA}} = \left\{ \sum_{h=1}^q N_{h\cdot\cdot} S_{ac,h} \right\}^2 / \left\{ \sum_{h=1}^q \frac{N_{h\cdot\cdot}^2}{(N_{h\cdot\cdot} - 1)} S_{a,h}^2 S_{c,h}^2 \right\} \quad (2.42)$$

Under H_0 , Q_{CMMA} asymptotically has the chi-square distribution with $d.f. = 1$. This test statistic is directed at the extent to which there is a consistent positive (or negative) association between the response scores and the sub-population scores in the respective tables. Thus, it is directed at 'average partial association' alternatives across the q tables between the response scores $\{a_h\}$ and the sub-population scores $\{c_h\}$. In particular, for the case where all the scores are either 0 or 1, Q_{CMMA} in (2.42) simplifies to the Cochran/Mantel-Haenszel statistic in (2.16) which corresponds to the collapsed set of (2×2) tables which are obtained by pooling the categories which are assigned '0' and those which are assigned '1' separately for the response categories and the sub-populations. Finally, if marginal rank or rdit-type scores are obtained from both the rows and columns of each table with mid-ranks assigned for ties, the statistic Q_{CMMA} in (2.42) is essentially equivalent to a partial Spearman rank correlation test, conditioning on the levels of the covariable set. Otherwise, the same types of comments could be made about power properties and central limit theory considerations for these statistics as were made for the multivariate test in section 2.1

and the mean score test in section 2.2, except that the concept of power is stronger relative to the association structure at which they are targeted and sample size requirements are minimal because only one degree of freedom is involved.

2.4. Summary

The general CMH procedure represents a flexible approach for investigating average partial association from several points of view. However, since the multiple hypergeometric model is only induced through the consideration of the basic hypothesis H_0 , other types of statistical analysis like the estimation of parameters and standard errors and the construction of confidence intervals cannot be undertaken in this framework because once the hypothesis H_0 is rejected the multiple hypergeometric model no longer is applicable. Thus, an alternative framework such as the fitting of log-linear models is necessary for such purposes. However, the test of H_0 in (2.1) within the log-linear model framework as discussed in Bishop, Fienberg and Holland [8] requires that the model for 'no second-order interaction' fits the data.

In contrast, the concept of 'no second-order interaction' is not a direct assumption of the CMH procedures because it is incorporated within the hypothesis being tested. Thus, such interaction is within the scope of the alternatives against which this method is directed, provided that such alternatives have a non-null average component. It should be noted that this concept involves the testing of a broad hypothesis by a method targeted at a narrow alternative. In other words, the Cochran/Mantel-Haenszel method is concerned with the broad hypothesis of no partial association and no interaction simultaneously. Thus, it can be rejected if contradicted in the direction of some type of narrow alternative. Thus, if the broad hypothesis is not true in an average sense because certain quantities exceed (or are exceeded by) their expectations in the respective tables, then the Cochran/Mantel-Haenszel test statistic will have power to detect it when targeted accordingly. Moreover, this statement applies whether or not such alternatives to H_0 involve interaction. Thus, the Cochran/Mantel-Haenszel procedure is valid in the presence of second-order interaction because the absence of such second-order interaction is incorporated within the overall

hypothesis which is being tested, and this implies that the presence of such interaction is within the scope of the alternatives against which the hypothesis is being tested, provided that such alternatives have a non-null average component.

However, the presence of interaction does tend to cause a loss of power in detecting partial association since some alternative other than the average is involved for such situations. Aside from this limitation, Birch [3,9] has indicated that if the hypothesis of ‘no second-order interaction’ does hold, then the Mantel-Haenszel test and its various appropriate generalizations have certain optimal test properties. Furthermore, because of the sample size requirements for testing the ‘no interaction’ hypothesis, the Cochran/Mantel-Haenszel strategy and its various appropriate generalizations is essentially the most flexible strategy available for testing for ‘average partial association’ when there are small sample sizes in many of the cells.

3. Specification of scores

The tests for average partial association in terms of mean scores discussed in section 2.2 and in terms of rank correlation coefficients in section 2.3 all require the specification of appropriate scores for the ordinally scaled variables. The choice of a particular set of scores depends on a variety of substantive and statistical issues which will not be elaborated further here. For such considerations, the reader is referred to Yates [10], Williams [11], Mantel [5], Bross [12], Bhapkar [13] and Koch et al. [14]. In most applications, there are essentially six basic types of scores which may be of interest. Each of these choices will be discussed in the following sections.

3.1. Natural scores

For some situations, the levels of an ordinal variable may represent well-defined intervals of an underlying quantitative variable such as age. Thus, a natural choice of scores may be the corresponding midpoints of the intervals. Otherwise, the specific physical or biological phenomenon may indicate a natural choice of scores. In such cases, H_0 in (2.1) can be investigated under these natural scores by choosing the $\{a_h\}$

and/or the $\{c_h\}$ to reflect these substantive considerations.

3.2. Binary partition scores

If the major question of interest is focused on the comparison of a particular combination of the levels of the response variable with the remaining levels, this implied binary split of the response variable can be investigated within the scope of mean scores as outlined in section 2.2 by assigning either 0 or 1 to each of the elements of the $\{a_h\}$. In particular, let $R = \{r_1, r_2, \dots, r_k\}$ denote the set of the k levels to be selected. Then for $h = 1, 2, \dots, q$ let:

$$a_{hj} = \begin{cases} 1, & \text{if } j \in R \\ 0, & \text{otherwise} \end{cases} \quad (3.1)$$

This choice of scores simply collapses the r levels of the response variable into a dichotomous variable with one category determined by the union of the k levels in R and the other category determined by the union of the remaining $(r - k)$ levels.

Furthermore, if one set of sub-populations is to be compared with the remaining set, this implied binary split among the sub-populations can also be created by the appropriate choice of the $\{c_h\}$. Specifically, let $S = \{s_1, s_2, \dots, s_\ell\}$ denote the set of sub-populations to be selected. Then for $h = 1, 2, \dots, q$ let:

$$c_{hi} = \begin{cases} 1, & \text{if } i \in S \\ 0, & \text{otherwise} \end{cases} \quad (3.2)$$

This choice of scores in (3.2) simply collapses the ℓ selected sub-populations into one group and the remaining $(s - \ell)$ sub-populations into the other group.

3.3. Uniform scores

As outlined in section 3.2, the binary partition scores do not require that the dimensions of the contingency table represent underlying continuous variables. However, if the response variable is ordinally scaled with progressively larger intensities, uniform (or equal-increment) scores specified for $h = 1, 2, \dots, q$ by letting:

$$a_{hj} = j \quad (3.3)$$

for $j = 1, 2, \dots, r$ can increase the statistical power in detecting differences among the sub-populations for progressive effects in the response variable. Similarly, if the sub-populations are ordinally scaled, uniform (or equal-increment) scores defined by:

$$c_{hi} = i \quad (3.4)$$

for $i = 1, 2, \dots, s$ and $h = 1, 2, \dots, q$ can be assigned. These scores in (3.3) and (3.4) can be obtained alternatively as sums of successive binary partitions of the corresponding levels of the response variable (or sub-populations) as outlined in Koch et al. [14]. Thus, these scores do not necessarily imply that the actual levels of the column (or row) variables are equally spaced in some sense.

3.4. Marginal rank scores

The analysis of contingency table data involving ordinally scaled variables can also be approached from the point of view of various non-parametric rank procedures. In particular, for the h -th level of the covariable set, all subjects who are jointly classified as belonging to the i -th sub-population and the j -th response category can be assumed to be tied for the corresponding levels of each variable. Consequently, by assigning mid-ranks to tied observations, the rank for the j -th level of the response variable in the h -th table is obtained as:

$$R_{h*j} = \{[(N_{h*j} + 1)/2] + \sum_{k=1}^{j-1} N_{h*k}\} \quad (3.5)$$

These ranks can then be utilized as the set of scores $\mathbf{a}'_h = (R_{h*1}, R_{h*2}, \dots, R_{h*r})$ for the h -th table. For this choice of scores in (3.5), it can be shown that the mean score procedures outlined in section 2.2 (case II) correspond to the Kruskal-Wallis method of ANOVA for each of the h tables. Thus, the Generalized Cochran/Mantel-Haenszel statistic Q_{CMMS} in (2.29) using the scores in (3.5) can be viewed as essentially equivalent to a partial Kruskal-Wallis ANOVA on ranks, conditioning on the levels of the covariable set.

If the sub-populations are also ordinally scaled, the rank for the i -th sub-population in the h -th table can be denoted by:

$$R_{hi*} = \{[(N_{hi*} + 1)/2] + \sum_{k=1}^{i-1} N_{hk*}\} \quad (3.6)$$

These ranks can then be utilized as the set of scores $\mathbf{c}'_h = (R_{h1*}, R_{h2*}, \dots, R_{hs*})$ for the h -th table. Using these rank scores in (3.5) and (3.6), it can be shown that the composite mean score procedures outlined in section 2.3 (case III) correspond to a Spearman rank correlation analysis for each of the h tables. Thus, the Generalized Cochran/Mantel-Haenszel statistic Q_{CMMMA} in (2.42) using the joint set of ranks from (3.5) and (3.6) can be viewed as essentially equivalent to a partial Spearman rank correlation test, conditioning on the levels of the covariable set.

3.5. Marginal rident-type scores

An alternative set of scores based on ranks can be proposed by expressing the ranks obtained in section 3.4 relative to the total sample size in the corresponding table. Specifically, for $h = 1, 2, \dots, q, j = 1, 2, \dots, r$ and $i = 1, 2, \dots, s$, let:

$$U_{h*j} = R_{h*j}/N_{h*} \quad (3.7)$$

$$U_{hi*} = R_{hi*}/N_{h*} \quad (3.8)$$

where the $\{R_{h*j}\}$ and $\{R_{hi*}\}$ are as defined in (3.5) and (3.6). These scores in (3.7) and (3.8) closely resemble the rident scores due to Bross [12] and have also been proposed by Mantel [5]. The basic difference between these rident-type scores and those outlined in Bross [12] is that the scores in (3.7) and (3.8) are obtained from the observed data in each of the h tables, as opposed to being derived from a relatively identified distribution (reference population).

The test statistics for partial association in case II and case III using the rident-type scores in (3.7) and (3.8) will be equivalent, but not necessarily identical, to those under the corresponding rank scores in (3.5) and (3.6). Furthermore, under case II these scores give rise to average ridents for each row of a given table and can be interpreted as probabilities that randomly selected individuals from the respective sub-population for a particular level of the covariable set have more extreme responses than a corresponding randomly selected subject from the marginal reference population of that table.

3.6. Combined rident-type scores

As discussed in the previous two sections, both the rank and rident-type scores are computed separately

for each of the q tables. This choice of scores has the advantage that the test statistics such as (2.24) and (2.37) utilize the marginal distributions of each table. Moreover, the Generalized Cochran/Mantel-Haenszel statistics such as (2.29) and (2.42) are also based on the specified margins of each table; and thus they represent appropriate tests for partial association.

On the other hand, it may also be of interest to perform these tests using the same set of ridit-type scores for each table. One approach to selecting such scores is to consider the combined data set obtained by summing the corresponding marginal distributions across the q tables. In particular, let $N_{.i} = \sum_h N_{hi}$ denote the total number of subjects classified as belonging to the i -th sub-population and let $N_{.j} = \sum_h N_{hj}$ denote the total number of subjects classified as belonging to the j -th response category. Then the overall rank for the j -th level of the response variable can be expressed as:

$$R_{**j} = \{[(N_{.j} + 1)/2] + \sum_{k=1}^{j-1} N_{.k}\} \quad (3.9)$$

Similarly, the overall rank for the i -th sub-population is obtained as:

$$R_{*i*} = \{[(N_{.i} + 1)/2] + \sum_{k=1}^{i-1} N_{.k}\} \quad (3.10)$$

From these quantities a set of overall ridit-type scores can be proposed by expressing these ranks in (3.9) and (3.10) relative to the overall total sample size N . Specifically, for $j = 1, 2, \dots, r$ and $i = 1, 2, \dots, s$, let:

$$U_{**j} = R_{**j}/N \quad (3.11)$$

$$U_{*i*} = R_{*i*}/N \quad (3.12)$$

Consequently, the tests for partial association outlined in sections 2.2 and 2.3 can be computed under the same set of overall ridit-type scores with the appropriate choices from either (3.11) and/or (3.12).

4. Data input

The data input options for PARCAT are quite flexible, permitting the data set to be read from any input device such as cards, tape or disk files. The data are read as a sequence of q : ($s \times r$) contingency tables

structured in the form of s rows and r columns as shown in table 1. Each of the q tables corresponds to one of the levels of the covariable set. The observed frequencies are handled in one of two different ways, depending on the size of $s * r$ as discussed in the following sections.

If the data are not already summarized in contingency table form, a general purpose statistical analysis package such as SAS described in Barr et al. [15] or SPSS discussed in Nie et al. [16] can be used to cross-classify the raw data into frequency counts according to the appropriate levels of the covariable set, the sub-populations, and the response profiles to conform to the structure indicated in table 1.

4.1. Frequency data: type 1

If $s * r \leq 100$, the observed frequencies in the format of table 1 (except for the row and column totals) are entered in row order for the tables indexed by $h = 1, 2, \dots, q$. Thus, there are $q * s$ records of data each containing r frequency counts.

4.2. Frequency data: type 2

If $s * r$ is in the range of 100–2500, the observed frequencies are still entered in row order for the tables indexed by $h = 1, 2, \dots, q$. However, the data need to satisfy the following two constraints:

- (i) $s \leq 100$;
- (ii) The measurement scale for the response variable must be at least ordinal so that a set of scores can be applied directly to the column categories to create a mean score function for each row of the table.

Under this type of data entry, the multivariate test statistics with $(s - 1)(r - 1)$ degrees of freedom under case I discussed in section 2.2 are not computed. Either case II or case III statistics are available as discussed in sections 2.3 and 2.4.

5. Use of PARCAT

The following sets of cards are used to enter the contingency table data and the parameters which determine the type of analysis to be performed:

- (0) JOB CONTROL CARDS

- (1) BASIC PARAMETER CARD
- (2) TITLE CARD
- (3) SCORE CARDS
- (4) DATA INPUT CARDS

Several analyses of a given data set can be performed in the same computer run by specifying different sets of scores. This can be accomplished by repeated use of as many sets of cards (1)–(3) as desired.

5.1. Description of the cards

Since the program is written in FORTRAN, all integer-valued parameters must be right-justified in their fields on the input cards. All FORMAT statements must be enclosed in parentheses and should be left-justified in their fields. In addition, the input data must be read according to floating-point specifications involving either F, E or G, e.g., 8F10.0, 6E13.5 or 10G8.0. Fixed-point specifications involving integer (I) format are not permissible.

(0) JOB CONTROL CARDS

These cards are necessary to access and to execute the load module of PARCAT. Because they will vary from one computer system to another, the user will need to determine the specific commands which are required at his/her computer installation. For example, under MTS at the University of Michigan, the required card is as follows:

```
$RUN SJS6:PARCAT 5=*SOURCE* 6=*SINK*
7=*SINK* 9=-T
```

Otherwise, a prototype of the required cards for a hypothetical IBM 360/370 installation is as follows:

```
//PARCAT JOB USER123 CLASS=F
//JOB LIB DD DSN=NAME,DISP=SHR,VOL=ABC
// EXEC PGM=PARCAT
//FT06F001 DD SYSOUT=A
//FT07F001 DD SYSOUT=A
//FT09F001 DD DSN=&&TEMP,DISP=NEW,
// UNIT=SYSDA,SPACE=(TRK(1,2)),
// DCB=(RECFM=VBS,BLKSIZE=3000)
//FT05F001 DD*
```

Further discussion of job control language cards is given in section 5.2 within the context of instructions for loading and executing the program.

(1) BASIC PARAMETER CARD

Columns	Information contained
5	Status of data set: 5 = new data; 6 = reanalysis of data saved from previous step
10	Type of input data (skip if columns 5 ≠ 5): 1 = $(s * r) \leq 100$; 2 = $s \leq 100$ and $100 \leq (s * r) \leq 2500$
14–15	Device number from which input data are to be read (skip if column 5 ≠ 5). [Do not use 06, 07, or 09]
18–20	Number of tables in data set (q)
21–22	Number of rows per table (s)
23–24	Number of columns per table (r)
26	Type of test statistics requested: (statistics for cases of lower numbering will automatically be provided): 1 = case I: Multivariate test (see section 2.1); 2 = case II: Mean score test (see section 2.2); 3 = case III: correlation test (see section 2.3)
28	Type of scores for columns
30	Type of scores for rows
	Type of scores available for coding in column 28 and/or 30: 1 = Binary partition scores (see section 3.2 and specify scores on (3) SCORE CARD); 2 = Uniform or equal increment scores (see section 3.3); 3 = Natural or user-specified scores (see section 3.1 and specify scores on a (3) SCORE CARD); 4 = Combined ridit-type scores (see section 3.6); 5 = Marginal rank scores (see section 3.4); 6 = Marginal ridit-type scores (see section 3.5)
31	Print option: 0 = Print individual table results and summary statistics; 1 = Print only summary statistics.

- 32 Save option:
0 = Do not save data for reanalysis;
1 = Save data for reanalysis
- 33–80 Format by which each row of the contingency tables will be read [Default = (8F10.0)]

(2) TITLE CARD

This card follows immediately after the (1) BASIC PARAMETER CARD, regardless of whether or not one or more (3) SCORE CARDS are required. The title can be any set of alphanumeric characters to describe the particular analysis specified by the preceding (1) BASIC PARAMETER CARD. This 80 character title will be printed at the head of the summary table. If no title is desired, a blank card must still be included for the (2) TITLE CARD.

- (3) SCORE CARD(S) [Skip if column 28 \neq 1 or 3 and column 30 \neq 1 or 3 of (1) BASIC PARAMETER CARD]

This card is used to enter either binary partition scores (type 1) or user supplied scores (type 3) for the columns and/or the rows of the contingency tables. If scores are requested for both columns and rows, i.e., column 28 and column 30 of the (1) BASIC PARAMETER CARD have either a 1 or 3, the SCORE CARD for columns must precede the SCORE CARD for rows. The format and score information is entered as follows, depending on whether the scores are type 1 or 3:

- (1) Enter the number of columns (or rows) in the specified binary partition followed by the column (or row) numbers included in the partition using the format specification (20I4);
- (3) Enter the user-specified scores according to the (10F8.0) format specification.

(4) DATA CARDS

Regardless of the input device, the contingency table frequencies are entered with each sub-population (row) beginning on a new record according to the format specified either by the default (10F8.0) or in columns 33–80 of the (1) BASIC PARAMETER CARD.

5.2. Instructions for loading and executing PARCAT

In order to load PARCAT on the computer system, it is necessary to compile, link-edit, and place the load module in a member of a partitioned data set. Usually such permanent data sets must be requested from installation personnel. Since the amount of space required for the partitioned data set depends on the storage device, the user is advised to seek aid from the appropriate installation personnel in determining the desired size.

The program is written in IBM 360/370 FORTRAN IV level G. The usual method of loading (on IBM equipment) is to use the IBM catalogued procedure FTGCL which utilizes the FORTRAN G compiler and linkage editor.

5.2.1. Sample compilation and loading

Suppose a partitioned data set called NAME has already been allocated and placed on disk volume ABC and that the source program for PARCAT is on a labelled tape volume NOTT1 on the first file called PARCAT. The load module can be placed in a member named PARCAT by using the following procedure:

```
//COMPILE JOB ACCT1.ABC,COOPER,T=(0,45),
MSGLEVEL=1,CLASS=B
// EXEC FTGCL
//FORT.SYSIN DD UNIT=TAPE VOL=NOTT1
DSN=PARCAT
//LKED.SYSLMOD DD DSN=NAME(PARCAT),
DISP=OLD,VOL=ABC
```

5.2.2. Assignment of I/O devices

- Unit 5 Input for parameter cards
- Unit 6 Output of individual tables and their associated statistics
- Unit 7 Output of summary statistics and CMH-statistics
- Unit 9 Used as temporary storage of data for reanalysis. The records are of variable length and are written and read with unformatted I/O statements. Recommended block size is 3000 bytes, since most record lengths will not exceed this order of magnitude. Record lengths could be as small as 104 bytes or as large as 40008 bytes if the largest possible problem was run. It is also important that rec-

ords be spanned across blocks; otherwise, a great deal of space may be wasted. Storage should be on any direct access unit the system chooses to assign.

Unit *I* The unit number *I* is specified in columns 14–15 on the (1) BASIC PARAMETER CARD to indicate the input unit from which the contingency table data will be read. This unit may be any permissible unit number except for the three reserved for output, viz. 06, 07, and 09. The user may specify *I*=5, in which case the data are read from the same card stream as the parameter cards. If a unit other than 5 is selected, a separate DD statement for the data file is required. For example, if 25 is specified in column 14–15 on the (1) BASIC PARAMETER CARD, the following DD statement must be included with the other JCL statements:

```
//FT25F001 DD DSN=DATA,DISP=OLD,
UNIT=DISK,VOL=ABC
```

for reading the contingency table data which are stored on a disk file called DATA.

5.2.3. Sample JCL for execution of PARCAT

Assuming that the user's input data are on cards

and that the load module has been stored as outlined in section 5.2.1, the following JCL statements illustrate a sample run of PARCAT:

```
//PARCAT JOB ACCT1.ABC,COOPER,
MSGLEVEL=1,T=2,CLASS=F
//JOB LIB DD DSN=NAME,DISP=SHR,VOL=ABC
// EXEC PGM=PARCAT
//FT06F001 DD SYSOUT=A,DCB=(RECFM=VBA,
LRECL=133,BLKSIZE=137)
//FT07F001 DD SYSOUT=A,DCB=(RECFM=VBA,
LRECL=133,BLKSIZE=137)
//FT09F001 DD DSN=&&TEMP,DISP=NEW,
UNIT=SYSDA,SPACE=(TRK,(1,2)),
DCB=(RECFM=VBS,BLKSIZE=3000)
//FT05F001 DD *,DCB=BLKSIZE=80
```

These cards are followed by the program and data cards as outlined in section 5.1.

6. Examples and sample input cards

This section presents two representative examples of the use of PARCAT with primary attention directed at the preparation of control cards discussed in section 5. More complete analyses of each of these data sets can be found in the papers cited in the corresponding sections.

Table 2
Deaths from leukemia observed at ABCC (1950–1970)

Age	Survival status ^a	Not in city	Dose (rads)				
			0–9	10–49	50–99	100–199	200+
0–9	LD	0	7	3	1	4	11
	NLD	5015	10752	2989	694	418	387
10–19	LD	5	4	6	1	3	6
	NLD	5973	11811	2620	771	792	820
20–34	LD	2	8	3	1	3	9
	NLD	5669	10828	2798	797	596	624
35–49	LD	3	19	4	2	1	10
	NLD	6158	12645	3566	972	694	608
50+	LD	3	7	3	2	2	6
	NLD	3695	9053	2415	655	393	289

^a LD denoted death from leukemia

NLD denotes non-death from leukemia

6.1. Leukemia mortality data

The following data in table 2 which appeared in Sugiura and Otake [4] show the numbers of deaths from leukemia (LD) observed at the Atomic Bomb Casualty Commission (ABCC) and the numbers of individuals who did not die from leukemia (NLD) during 1950–1970 according to age at the time of the atomic bomb and the estimated radiation dosage. Given the framework outlined in section 2, these data in table 2 involve $s = 6$ sub-populations (dose), $r = 2$ response categories (survival status) and $q = 5$ levels of the covariable (age at exposure). For convenience, the data are displayed in transposed form from the structure denoted in table 1.

The primary objective in the analysis of these data is to study historically a sample of persons who died from leukemia and to evaluate the extent to which death was related to the estimated radiation dosage adjusting for age at exposure. In a strict sense, all of the frequency counts in table 2 are fixed constants (as opposed to random variables) because of the historical nature of the data. On the other hand, there is still interest in the hypothetical question of whether or not the observed distribution of leukemia deaths is at random with respect to the estimated radiation dosage for each of the age groups. Consequently, the basic framework for this investigation is the multiple hypergeometric model in (2.2). For these data, the generalized CMH statistic in (2.13) yields $Q_{CMH} = 461.7$, which asymptotically has a chi-square distribution with $d.f. = 5$ under H_0 in (2.1). Equivalently, the mean score test in (2.29) under the scores $a'_h = (1, 0)$ for $h = 1, 2, \dots, 5$ also yields $Q_{CMMS} = 461.7$. This equality between Q_{CMH} and Q_{CMMS} results from the

fact that the response variable has only $r = 2$ levels and holds for any non-trivial choice of a_h .

Furthermore, since the levels of radiation dose have an underlying continuous distribution, various sets of scores can be applied to these categories to investigate order partial association alternatives in terms of correlation test statistics with $d.f. = 1$ as outlined in section 2.3. This approach is directed at the extent to which the probability of leukemia death increases (or decreases) with the increase of radiation dosage. In particular, four sets of scores discussed in section 3 are potentially of interest in this context. The first is uniform (or equal increment) scores which can be obtained as sums of successive binary partitions of the dosage scale which have been assigned equal weights. Thus, these scores do not necessarily imply that the actual levels of the categories are equally spaced in some sense. Another choice of scores involves ridsits computed from the first-order marginal distribution of leukemia deaths and radiation dosage levels obtained by summing across age groups. These scores are then applied to each of the subtables determined by the age groups to provide a partial Spearman rank correlation analysis of these data. Alternatively, a natural choice of scores is the midpoint of the radiation dosage intervals, where arbitrarily 0 is assigned to the category 'not in city' and 300 is assigned to the category '200+'. Moreover, for substantive reasons related to inverse square principles, it is of interest to investigate the increase of the probability of leukemia death with respect to the square root of the radiation dosage midpoint. These four sets of scores together with their corresponding correlation test statistic Q_{CMMA} obtained from (2.42), are displayed in table 3. Thus, we conclude

Table 3
Radiation dosage scores and correlation test statistics

Type of scores	Not in city	Radiation dose levels					Q_{CMMA}
		0–9	10–49	50–99	100–199	200+	
Uniform	0	1	2	3	4	5	262.6
Combined ridsits	0.126	0.512	0.841	0.928	0.960	0.987	102.4
Midpoints	0	5	30	75	150	300	426.3
Square root of midpoints	0	2.24	5.48	8.66	12.25	17.32	316.5

COLUMN	1	2	3	4	5	6
123456789012345678901234567890123456789012345678901234567890						
5	1	5	5 6 2 3 2	201(265.0)		
DEATHS FROM LEUKEMIA OBSERVED AT ABCC (1950-1970)						

Fig. 1.

that there is a highly significant ($\alpha = 0.01$) radiation dosage effect under any of these scores adjusting for the effects of age at exposure. A more extensive analysis of these data is reported in Landis et al. [7].

The card preparation necessary for analyzing these data by PARCAT is described in the following paragraphs.

(1) BASIC PARAMETER CARD

For each of the $q = 5$ tables corresponding to the age groups there are $r = 2$ responses within each of $s = 6$ sub-populations so that $s * r = 12 \leq 100$, which

permits the data to be entered according to input mode 1. The required parameter values to analyze these data under uniform scores for both the row and column dimensions are indicated in the appropriate columns of the first card in fig. 1.

(2) TITLE CARD

A title for this particular step of the analysis is listed as card 2 in fig. 1. Since these data will be processed under several sets of scores sequentially, it is necessary to set column 32 = 1 as indicated in fig. 1.

COLUMN	1	2	3	4	5	6
123456789012345678901234567890123456789012345678901234567890						
0.	5015.					
7.	10752.					
3.	2989.					
1.	694.					
4.	418.					
11.	387.					
5.	5973.					
4.	11811.					
6.	2620.					
1.	771.					
3.	792.					
6.	820.					
2.	5669.					
8.	10828.					
3.	2798.					
1.	797.					
3.	596.					
9.	624.					
3.	6158.					
19.	12645.					
4.	3566.					
2.	972.					
1.	694.					
10.	608.					
3.	3695.					
7.	9053.					
3.	2415.					
2.	655.					
2.	393.					
6.	289.					

Fig. 2.

(3) SCORE CARD

No score card is required since uniform scores are selected for both rows and columns (col. 28 = col. 30 = 2 on card 1) for this step.

(4) DATA CARDS

The data from table 2 are arrayed in the standard layout of table 1 as shown in fig. 2. Note that the column alignment conforms to the format specification on card 1.

Although the individual tables together with their corresponding test statistics are each printed, only the summary table will be shown in fig. 3 for illustrative purposes. Note that the individual table statistics from (2.9), (2.24) and (2.37) are given in section A of fig. 3, together with the appropriate degrees of free-

dom and the associated significance level of each chi-square variable. In addition, the corresponding sums of these quantities given in (2.10), (2.25) and (2.38), respectively, are given in the row labeled 'TOTAL'. Finally, the Generalized CMH statistics defined in (2.13), (2.29) and (2.42) are given in section B of the printed output.

(5) REANALYSIS OF DATA

These data can be reanalyzed under the additional sets of row scores outlined in table 3 by the sequence of control cards given in fig. 4. In particular, note that column 32 = 1 for the first control card in fig. 1 as well as each additional (1) BASIC PARAMETER CARD (column 5 = 6) to ensure that the same data set is saved on a temporary file for reanalysis.

```

***** PARCAT *****

GENERALIZED COCHRAN-MANTEL-HAENSZEL TEST STATISTICS
FOR AVERAGE PARTIAL ASSOCIATION IN THREE-WAY CONTINGENCY TABLES

DEATHS FROM LEUKEMIA OBSERVED AT ABCC (1950-1970)

COLUMN SCORES : UNIFORM          1.00  2.00
ROW SCORES   : UNIFORM          1.00  2.00  3.00  4.00  5.00  6.00

SUMMARY ACROSS TABLES

*****
A. SUMMARY OF INDIVIDUAL TABLE STATISTICS

TABLE NO.   SAMPLE SIZE      MULTIVARIATE Q      D.F.   P      MEAN SCORE QMS      D.F.   P      CORRELATION QMA      D.F.   P
1          20281      248.05      5      0.0      248.05      5      0.0      127.51      1      0.0
2          22812      43.89       5      0.0      43.89       5      0.0      29.32       1      0.0
3          21338      100.56      5      0.0      100.56      5      0.0      60.23       1      0.0
4          24682      88.85       5      0.0      88.85       5      0.0      38.02       1      0.0
5          16523      84.74       5      0.0      84.74       5      0.0      40.56       1      0.0
TOTAL     105636      566.09     25      0.0      566.09     25      0.0      295.65      5      0.0

B. GENERALIZED COCHRAN-MANTEL-HAENSZEL STATISTICS

SAMPLE SIZE      MULTIVARIATE Q(CMH)      D.F.   P      MEAN SCORE Q(CMMS)      D.F.   P      CORRELATION Q(CMMA)      D.F.   P
105636          461.66      5      0.0      461.66      5      0.0      262.62      1      0.0
*****

```

Fig. 3.

COLUMN	1	2	3	4	5	6
12345678901234567890123456789012345678901234567890	6	5	6	2	3	4
DEATHS FROM LEUKEMIA OBSERVED AT ABCC (1950-1970)	6	5	6	2	3	2
0.	5.	30.	75.	150.	300.	
DEATHS FROM LEUKEMIA OBSERVED AT ABCC (1950-1970)	6	5	6	2	3	2
0.	2.24	5.48	8.66	12.25	17.32	

Fig. 4.

6.2. Dumping syndrome data

The following data in table 4 which appeared in Grizzle, Starmer and Koch [17] were obtained from a multi-center randomized clinical trial involving suitably eligible patients who were treated in four participating hospitals during 1966–1968. Each patient was randomly assigned to one of the following four surgical procedures for the treatment of duodenal ulcers:

- A. Drainage and vagotomy;
- B. Antrectomy (25% resection) and vagotomy;
- C. Hemigastrectomy (50% resection) and vagotomy;
- D. 75% resection;

Table 4
Severity of dumping syndrome

Hospital	Treatment	Severity of syndrome ^a			Total
		N	S	M	
1	A	23	7	2	32
	B	23	10	5	38
	C	20	13	5	38
	D	24	10	6	40
2	A	18	6	1	25
	B	18	6	2	26
	C	13	13	2	28
	D	9	15	2	26
3	A	8	6	3	17
	B	12	4	4	20
	C	11	6	2	19
	D	7	7	4	18
4	A	12	9	1	22
	B	15	3	2	20
	C	14	8	3	25
	D	13	6	4	23

^a N denotes none; S denotes slight; M denotes moderate

and was subsequently classified with respect to the severity of the ‘dumping syndrome’ an undesirable sequela of surgery for duodenal ulcer.

The primary objective in the analysis of these data is to investigate the relationship between the nature and extent of surgery which involves the removal of different amounts of the stomach and the severity of the ‘dumping syndrome’ adjusting for potential hospital effects. From a strict randomization point of view, the basic framework for hypothesis testing is the multiple hypergeometric model in (2.2). Within the framework outlined in section 2, these data involve $s = 4$ sub-populations (treatments), $r = 3$ response categories (severity of dumping syndrome) and $q = 4$ levels of the covariable (hospital).

For these data, the Generalized CMH statistic in (2.13) is $Q_{CMH} = 10.60$ with $d.f. = 6$, which is not statistically significant ($\alpha = 0.10$). However, since the dumping syndrome severity is measured on an ordinal scale, the types of alternative hypotheses which are of primary interest are location shifts which are indicative of the extent to which the syndrome tends to be more severe for certain operations than others. Under uniform scores discussed in section 3.3, the mean score test in (2.29) is $Q_{CMMS} = 6.59$ with $d.f. = 3$ which is statistically significant ($\alpha = 0.10$). Alternatively, using marginal ridit-type scores as outlined in section 3.5, these data yield $Q_{CMMS} = 7.63$ with $d.f. = 3$ which is also statistically significant ($\alpha = 0.10$).

Finally, since the four surgical operations can be regarded as having an underlying ordinal scale with respect to the amount of stomach removed, it is also of interest to investigate the extent to which the location shifts in the severity of the dumping syndrome are related to the ordinal scale for treatment. Conse-

COLUMN	1	2	3	4	5	6
12345678901234567890123456789012345678901234567890						
5	1	5	4 4 3 3 2 2 1	(8G10.0)		
SEVERITY OF DUMPING SYNDROME						
23.	7.	2.				
23.	10.	5.				
20.	13.	5.				
24.	10.	6.				
18.	6.	1.				
18.	6.	2.				
13.	13.	2.				
9.	15.	2.				
8.	6.	3.				
12.	4.	4.				
11.	6.	2.				
7.	7.	4.				
12.	9.	1.				
15.	3.	2.				
14.	8.	3.				
13.	6.	4.				
6		6 6				
SEVERITY OF DUMPING SYNDROME						

Fig. 5.

***** PARCAT *****

GENERALIZED COCHRAN-MANTEL-HAENSZEL TEST STATISTICS
FOR AVERAGE PARTIAL ASSOCIATION IN THREE-WAY CONTINGENCY TABLES

SEVERITY OF DUMPING SYNDROME

COLUMN SCORES :	UNIFORM	1.00	2.00	3.00						
ROW SCORES :	UNIFORM	1.00	2.00	3.00	4.00					

SUMMARY ACROSS TABLES

A. SUMMARY OF INDIVIDUAL TABLE STATISTICS

TABLE NO.	SAMPLE SIZE	MULTIVARIATE			MEAN SCORE			CORRELATION		
		Q	D.F.	P	QMS	D.F.	P	QMA	D.F.	P
1	148	3.40	6	0.7575	2.62	3	0.4536	1.57	1	0.2105
2	105	10.82	6	0.0942	7.34	3	0.0617	7.06	1	0.0079
3	74	3.08	6	0.7987	1.69	3	0.6382	0.16	1	0.6927
4	90	5.21	6	0.5177	1.68	3	0.6420	0.66	1	0.4161
TOTAL	417	22.50	24	0.5494	13.34	12	0.3450	9.45	4	0.0508

B. GENERALIZED COCHRAN-MANTEL-HAENSZEL STATISTICS

SAMPLE SIZE	MULTIVARIATE			MEAN SCORE			CORRELATION		
	Q(CMH)	D.F.	P	Q(CMMS)	D.F.	P	Q(CMMA)	D.F.	P
417	10.60	6	0.1016	6.59	3	0.0862	6.34	1	0.0118

Fig. 6.

***** PARCAT *****
 GENERALIZED COCHRAN-MANTEL-HAENSZEL TEST STATISTICS
 FOR AVERAGE PARTIAL ASSOCIATION IN THREE-WAY CONTINGENCY TABLES

SEVERITY OF DUMPING SYNDROME

COLUMN SCORES : MARGINAL RIDITS SCORES VARY FOR EACH TABLE

ROW SCORES : MARGINAL RIDITS SCORES VARY FOR EACH TABLE

SUMMARY ACROSS TABLES

A. SUMMARY OF INDIVIDUAL TABLE STATISTICS

TABLE NO.	SAMPLE SIZE	MULTIVARIATE			MEAN SCORE			CORRELATION		
		Q	D.F.	P	QMS	D.F.	P	QMA	D.F.	P
1	148	3.40	6	0.7575	2.75	3	0.4312	1.41	1	0.2350
2	105	10.82	6	0.0942	8.93	3	0.0303	8.49	1	0.0036
3	74	3.08	6	0.7987	1.87	3	0.6004	0.20	1	0.6568
4	90	5.21	6	0.5177	1.96	3	0.5809	0.35	1	0.5523
TOTAL	417	22.50	24	0.5494	15.51	12	0.2149	10.46	4	0.0334

B. GENERALIZED COCHRAN-MANTEL-HAENSZEL STATISTICS

SAMPLE SIZE	MULTIVARIATE			MEAN SCORE			CORRELATION		
	Q (CMH)	D.F.	P	Q (CMMS)	D.F.	P	Q (CMMA)	D.F.	P
417	10.60	6	0.1016	7.63	3	0.0542	6.92	1	0.0085

Fig. 7.

quently, the correlation mode of the CMH procedure outlined in section 2.3 can be utilized to examine the extent to which the probability of more severe dumping syndrome outcomes tends to increase (or decrease) with the extent of the operation in terms of larger amounts of stomach removed. Specifically, the correlation test in (2.42) yields $Q_{CMMA} = 6.34$ for uniform scores and $Q_{CMMA} = 6.92$ for marginal ridit-type scores. Under H_0 in (2.1), these statistics asymptotically have chi-square distributions with $d.f. = 1$. The former is significant at $(\alpha = 0.05)$ and the latter at $(\alpha = 0.01)$. Thus, both indicate a significant relationship between dumping syndrome severity and the extent of operation.

The control cards for PARCAT necessary for analyzing these data under both uniform and marginal ridit-type scores are shown in fig. 5 and the corresponding summary tables of test statistics for each of these scoring procedures are shown in fig. 6 and 7, respectively.

7. Hardware specifications

This computer program is written in double precision in IBM 360/370 FORTRAN IV which incorporates a few extensions to American National Standard (ANS) FORTRAN. As a result, minor modifications of the source code may be required in order to use this program on other machines.

8. Program availability

The source deck for the batch version of PARCAT, together with the corresponding listings and test data sets from section 6, may be obtained for a nominal cost from the Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, MI 48109, USA. A current version of the documentation and running instructions is included with the initial purchase of the computer program.

9. Disclaimer

Although PARCAT has received extensive testing, no warranty, expressed or implied, is made to the accuracy and functioning of the program. No responsibility is assumed by the authors. However, if specific questions or problems do arise, contact the first author at the Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, MI 48109, USA.

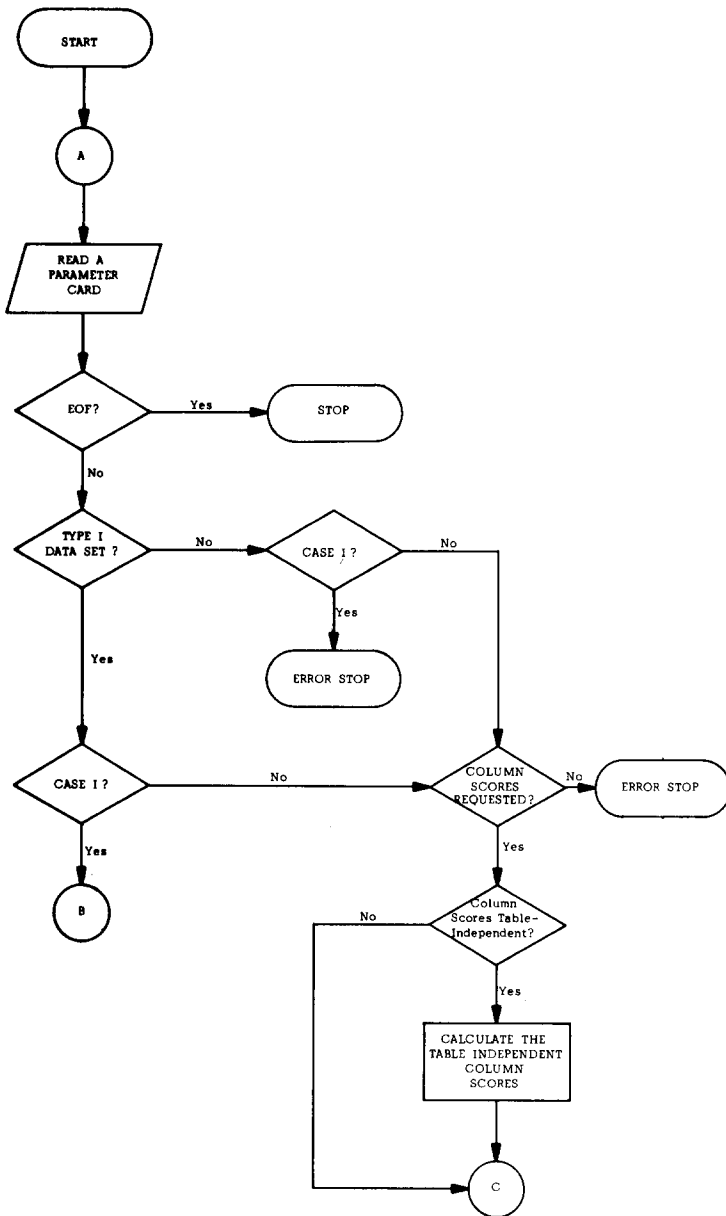
Acknowledgements

The authors wish to thank numerous graduate students in the Department of Biostatistics at the University of Michigan who provided many helpful comments during the debugging of various test versions of PARCAT. In particular, Martha Alway, James Lepkowski, Michael J. Long, and Janice Valentine were helpful in numerous test runs on large data sets which isolated various problems in the preliminary versions of the program. Moreover, the authors are grateful to David F. Hunsche who prepared the final form of the flowcharts shown in the Appendix and to Ms Connie Massey and Ms Rebecca Turner for the conscientious typing of the manuscript.

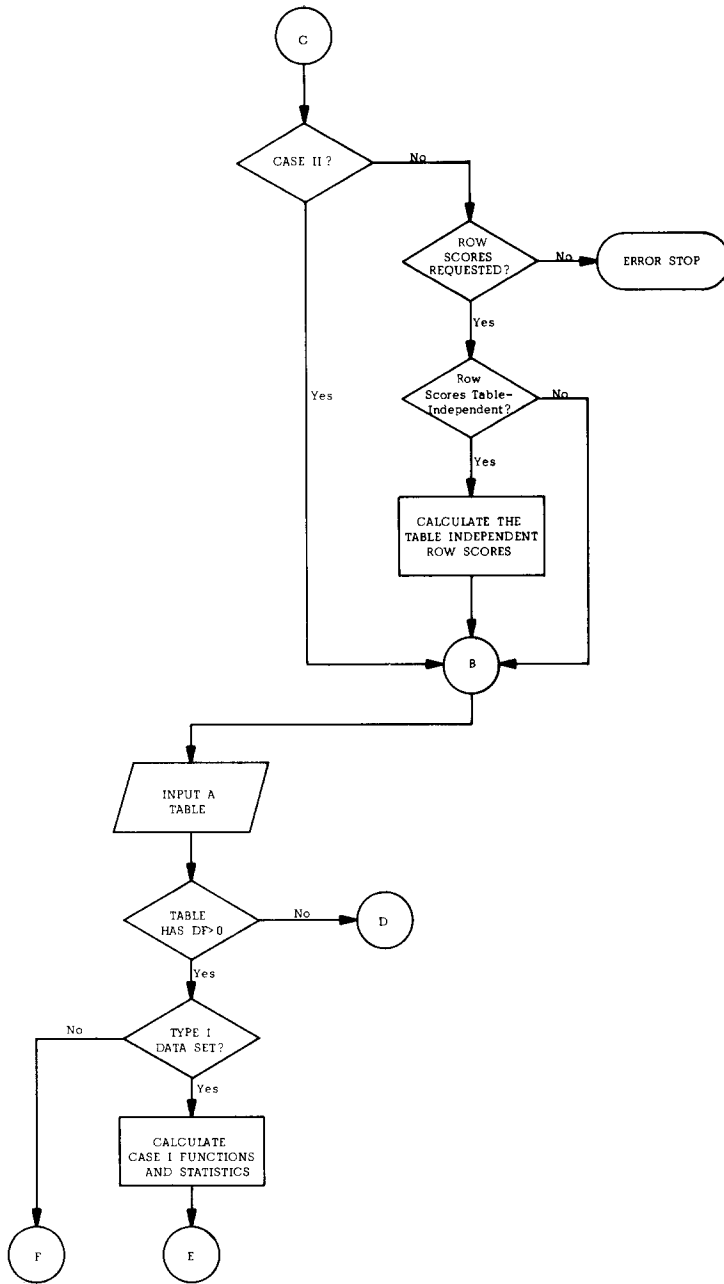
References

- [1] W.G. Cochran, Some methods for strengthening the common χ^2 test, *Biometrics* 10 (1954) 417–451.
- [2] N. Mantel and W. Haenszel, Statistical aspects of the analysis of data from retrospective studies of disease, *J. Nat. Cancer Inst.* 22 (1959) 719–748.
- [3] M.W. Birch, The detection of partial association, II: the general case, *J. Roy. Stat. Soc. B* 27, (1965) 111–124.
- [4] N. Sugiura and M. Otake, An extension of the Mantel-Haenszel procedure to $K \times 2 \times C$ contingency tables and the relation to the logit model, *Commun. Stat.* 3/9 (1979) 829–842.
- [5] N. Mantel, Chi-square tests with one degree of freedom: extensions of the Mantel-Haenszel procedure, *J. Am. Stat. Assoc.* 58 (1963) 690–700.
- [6] G.G. Koch and D.W. Reinfurt, An analysis of the relationship between driver injury and vehicle age for automobiles involved in North Carolina accidents during 1966–1970, *Accid. Anal. Prev.* 6 (1974) 1–18.
- [7] J.R. Landis, E.R. Heyman and G.G. Koch, Average partial association in three-way contingency tables: a review and discussion of alternative tests, *Int. Stat. Rev.* (1978) 237–254.
- [8] Y.M.M. Biship, S.E. Fienberg and P.W. Holland, *Discrete Multivariate Analysis* (The MIT Press, Cambridge, 1975).
- [9] M.W. Birch, The detection of partial association, I: the 2×2 case, *J. Roy. Stat. Soc. B* 26 (1964) 313–324.
- [10] F. Yates, The analysis of contingency tables with groupings based on quantitative characters, *Biometrika* 35 (1948) 176–181.
- [11] E.J. Williams, Uses of scores for the analysis of association in contingency tables, *Biometrika* 39 (1952) 274–289.
- [12] I.D.J. Bross, How to use ‘ridit’ analysis, *Biometrics* 14 (1958) 18–38.
- [13] V.P. Bhapkar, On the analysis of contingency tables with a quantitative response, *Biometrics* 24 (1968) 329–338.
- [14] G.G. Koch, J.R. Landis, J.L. Freeman, D.H. Freeman, jr and R.G. Lehnen, A general methodology for the analysis of experiments with repeated measurement of categorical data, *Biometrics* 33 (1977) 133–158.
- [15] A.J. Barr, J.H. Goodnight, J.P. Sall and J.T. Helwig, *A User’s Guide to SAS* (SAS Inst., 1976).
- [16] N. Nie, C.H. Hull, J.G. Jenkins, K. Stembrenner and D. Brent, *Statistical Package for the Social Sciences* (McGraw Hill, 1975).
- [17] J.E. Grizzle, C.F. Starmer and G.G. Koch, Analysis of categorical data by linear models, *Biometrics* 25 (1969) 489–504.

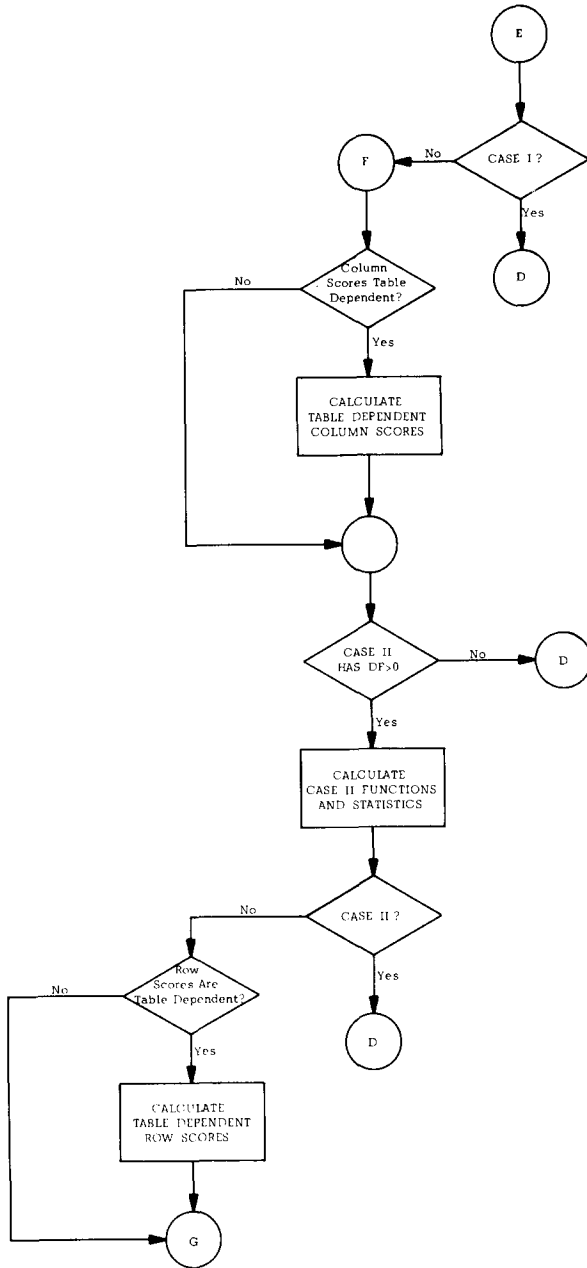
Appendix



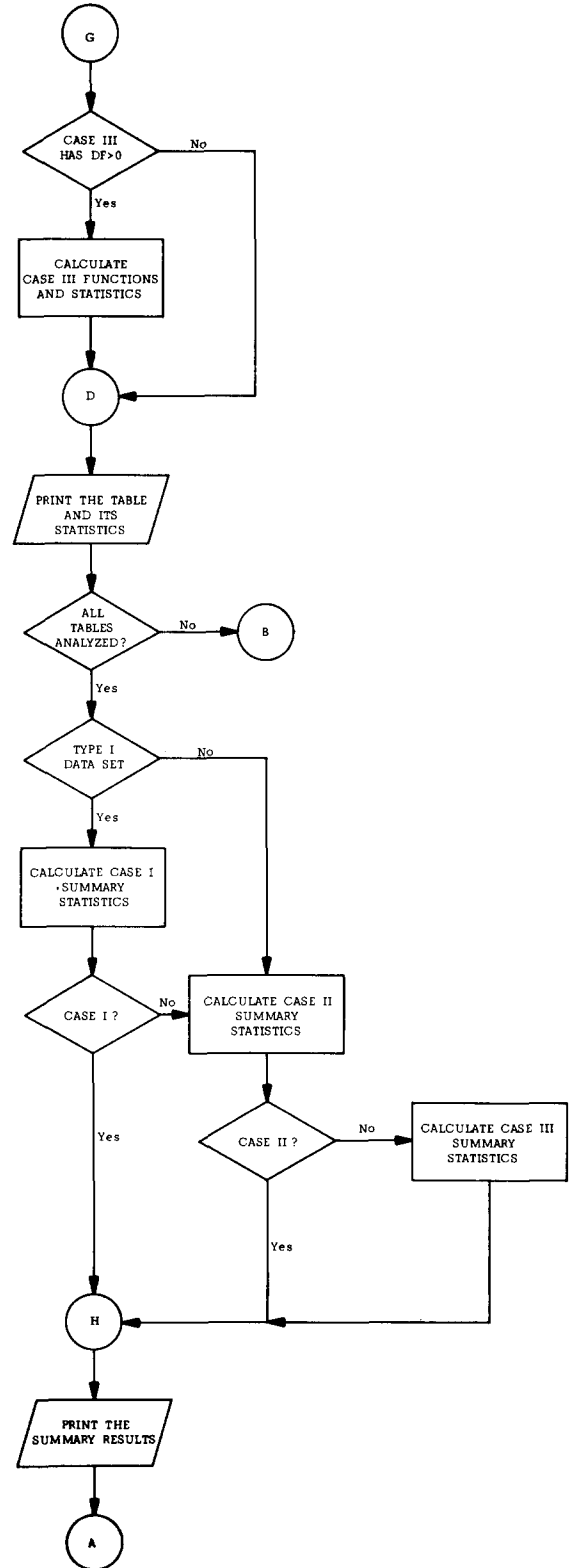
Flowchart 1.



Flowchart 2.



Flowchart 3.



Flowchart 4.