The Annals of Statistics
1985, Vol. 13, No. 1, 10–69

THE 1983 HENRY L. RIETZ MEMORIAL LECTURE[1]

THE ANALYSIS OF CROSS-CLASSIFIED DATA HAVING ORDERED AND/OR UNORDERED CATEGORIES: ASSOCIATION MODELS, CORRELATION MODELS, AND ASYMMETRY MODELS FOR CONTINGENCY TABLES WITH OR WITHOUT MISSING ENTRIES

BY LEO A. GOODMAN[2]

University of Chicago

CONTENTS

1. Introduction and summary. Let me begin this lecture in commemoration of Henry L. Rietz, the first president of the Institute of Mathematical

Received July 1984; revised October 1984.

[1] This paper is the written version of the Henry L. Rietz Memorial Lecture for 1983, presented at the invitation of the Institute of Mathematical Statistics, at the joint annual meeting of the Institute of Mathematical Statistics, the American Statistical Association, the Canadian Statistical Society, and the Biometric Society, in Toronto, Canada, on 16 August 1983. Part of the research presented here (the material on the relationship between association models and correspondence analysis) was discussed briefly by the author at the International Meeting on the Analysis of Multidimensional Contingency Tables, at the University of Rome, under the auspices of the Italian Statistical Society, on 26 June 1981.

[2] Support for this research was provided in part by National Science Foundation Grant SES-8303838. The written version of the Rietz Memorial Lecture was completed while the author was a

Statistics, by noting that his main research in statistics was on the theory and measurement of correlation (Rietz and Shade, 1908; Rietz and Smith, 1910; Rietz, 1912, 1916, 1919, 1920, 1924 (Chapter 8), 1927 (Chapter 4), 1932); and that a major part of my lecture will be devoted to the further development of this subject in situations where the variables under study pertain to qualitative/categorical classifications.

Henry Rietz began his career in statistics seventy-five years ago with his correlation studies, and he continued to contribute to this topic (among others) for roughly the next twenty-five years. In addition to his research work, Rietz took great pride in his teaching; and among his students we find Samuel S. Wilks, who also began his carrer in statistics (more than fifty years ago) with a strong interest in the study of correlation (see Wilks, 1931, and related articles cited in, e.g., Wilks, 1962). In view of Rietz's teacher–student connection with Wilks, and Wilks' teacher–student connection with me, the audience may view this Rietz Memorial Lecture as transmitted, in a sense, by lineal descent.

In the past, correlation (i.e. the usual Pearsonian product-moment correlation coefficient) has not played an important role in my work on the association between qualitative variables. Fifteen years ago, when the development of log-linear models for the analysis of qualitative variables was in its beginning stages, my R. A. Fisher Memorial Lecture made no mention of correlation (see Goodman, 1968); and thirty years ago, when joint work with William Kruskal on measures of association was in its beginning stages, there was only a passing reference to correlation (see Goodman and Kruskal, 1954). (In Kruskal's 1958 article on ordinal measures of association, there was some discussion of this topic, "for the sake of completeness and contrast.") Subsequent work on log-linear models and on measures of association continued either to ignore this topic or to pay only minimal attention to it (see, e.g., Goodman, 1978; and Goodman and Kruskal, 1979). However, with the introduction of association models for the analysis of qualitative variables (as described in, e.g., Goodman, 1979a), and the discovery that results obtained with these models are often related to (but different from) corresponding results obtained with canonical correlation analysis, interest grew in the study of correlation methods for the analysis of qualitative variables (see, e.g., Goodman, 1981a). In this Rietz Memorial Lecture, these methods will be developed further.

To illustrate the models and methods that will be discussed in this lecture, I shall reconsider three cross-classification tables analyzed in the earlier statistical

---

Fellow at the Center for Advanced Study in the Behavioral Sciences, with financial support provided in part by National Science Foundation Grant BNS-8011494 to the Center. For helpful comments, the author is indebted to S. J. Haberman, O. D. Duncan, C. C. Clogg, Z. Gilula, W. H. Kruskal, R. L. Plackett, A. Stuart, and D. L. Wallace. For helpful programming, the author is indebted to R. Kuziel and P. Norton.

AMS 1980 *subject classifications.* 62H17, 62H20, 62-07.

Key words and phrases. Contingency tables, association models, canonical correlation, correspondence analysis, ordinal and non-ordinal categorical data, log-linear and non- log-linear models, quasi-symmetry and generalized symmetry models, order restrictions.

literature: Table 1, studied earlier by Treloar (1939, page 228), Williams (1952), and Kendall and Stuart (1979, page 619); Table 2, studied earlier by Srole et al. (1962, page 213), Haberman (1974a; 1979, page 375), Gross (1981), Escoufier (1982), and Gilula and Haberman (1984); Table 3, considered earlier by many statisticians, including Stuart (1953, 1955), Tukey (1977, page 519), Gokhale and Kullback (1978, page 255), Kendall and Stuart (1979, page 618), Plackett (1981, page 25), and Agresti (1984, page 215).

Many different kinds of cross-classification tables can be analyzed using the models and methods that will be discussed in this lecture. I use these three examples to facilitate comparison with the methods discussed in the earlier literature and to illustrate a variety of problems that can arise in analyzing such

TABLE 1

Cross-classification of 135 *women according to their periodontal condition and calcium intake level*

periodontal condition	calcium intake level				Total
	1	2	3	4	
A	5	3	10	11	29
B	4	5	8	6	23
C	26	11	3	6	46
D	23	11	1	2	37
Total	58	30	22	25	135

TABLE 2

Cross-classification of 1,660 *people according to their mental health and their parents' socioeconomic status*

mental health status	parents' socioeconomic status						Total
	A	B	C	D	E	F	
well	64	57	57	72	36	21	307
mild symptom formation	94	94	105	141	97	71	602
moderate symptom formation	58	54	65	77	54	54	362
impaired	46	40	60	94	78	71	389
Total	262	245	287	384	265	217	1660

TABLE 3

Cross-classification of 7,477 *women according to their right eye grade and left eye grade with respect to unaided distance vision*

right eye grade	left eye grade				Total
	Best (1)	Second (2)	Third (3)	Worst (4)	
Best (1)	1520	266	124	66	1976
Second (2)	234	1512	432	78	2256
Third (3)	117	362	1772	205	2456
Worst (4)	36	82	179	492	789
Total	1907	2222	2507	841	7477

data and in interpreting the results that can be obtained by applying these models and methods.

The class of models presented in this lecture includes models that can be applied to $I \times J$ cross-classification tables in which the row categories and/or column categories are considered to be either unordered or ordered, and in which the spacing (or distance) between the ordered categories is considered to be either unspecified or specified. I shall not distinguish here between the case where the categories of a classification (either the row classification or the column classification) are considered unordered (e.g., categories pertaining to a classification that is nominal in principle) and the case where the categories are considered ordered but the ordering is unspecified (e.g., categories pertaining to a classification that is ordinal in principle but the ordering of the particular categories is not known *a priori*). In my initial article on association models (Goodman, 1979a), I considered a class of models that included models for the various kinds of cross-classification tables just described. This class of models (which I shall describe more fully in Section 2.1) consists of the null association model (the 0 model), the uniform association model (the U model), the row-effects association model (the R model), the column-effects association model (the C model), and two association models that included both row effects and column effects (the RC model and the R + C model). (The RC and R + C models were called models II and I, respectively, in Goodman, 1979a.) Table 4A describes the various cross-

TABLE 4A

Models for the scores of the row categories and the column categories in a cross-classification table, with unspecified or specified order for the categories, and with unspecified or specified spacing between the categories

models	unspecified order	specified order	
		unspecified spacing	specified spacing
0	Rows and Columns		
U			Rows and Columns
R	Rows	Rows*	Columns
C	Columns	Columns*	Rows
RC	Rows and/or Columns	Columns and/or Rows*	
R + C			Rows and Columns

* With specified order and unspecified spacing for the categories, the R, C, and RC models are modified as indicated later herein.

TABLE 4B

Degrees of freedom for the models applied to the $I \times J$ table

models	degrees of freedom
0	$(I-1)(J-1)$
U	$IJ-I-J$
R	$(I-1)(J-2)$
C	$(I-2)(J-1)$
RC	$(I-2)(J-2)$
R + C	$(I-2)(J-2)$

classification tables and the kind of association model suited to each kind of table. For example, the null association model (the 0 model), which is the usual model of statistical independence between the row classification and the column classification, is suited to the case where the order of the row categories and the order of the column categories is unspecified (i.e., the model is invariant with respect to permutation of the rows and permutation of the columns, and so information about the order of the row catgories and the order of the column categories is not used in the model); while the uniform association model (the U model) is suited to the case where both the order of the row categories and the order of the column categories are specified, and both the spacing between the ordered row categories and the spacing between the ordered column categories are also specified. Of course, a model that is suited to the case in which a classification's categories have, say, unspecified order can also be applied when the classification's categories have specified order; but the information about the specified order is not used in the model. Similarly, a model that is suited to the case in which a classification's categories have, say, specified order with unspecified spacing between the ordered categories can also be applied when the classification's categories have specified order with specified spacing between the ordered categories; but the information about the specified spacing is not used in the model.

When the order of the row categories and the order of the column categories are specified, use of the natural numbers (i.e., the integers) has been proposed as a simple scoring system for the ordered row categories and the ordered column categories (see, e.g., Yates, 1948); and other relatively simple scoring systems have also been considered (see, e.g., Kendall and Stuart, 1979, page 597). However, suitable statistical methods have not been available for determining whether a proposed scoring system yields an adequate picture of the association between the row classification and the column classification in a given $I \times J$ cross-classification table. If the scoring system is intended to yield a picture of the association in the $I \times J$ table (with association measured in the way described in Section 2.1 below), then the class of association models considered in my earlier article (Goodman, 1979a) can be used to provide a statistical test of whether the proposed scoring system yields an adequate picture. On the other hand, if instead of measuring association as we do with association models (and as we do in Section 2.1 below), we use a different measure of "nonindependence," then a different class of models is needed. To meet this need, I shall consider in Section 2.2 a different class of models, which I shall call correlation models (for reasons that will become clear later). If the scoring system is intended to yield a picture of the nonindependence in the $I \times J$ table (with nonindependence measured in terms of the Pearsonian correlation beween unknown true row scores and unknown true column scores), then the class of correlation models considered in this lecture can be used to provide a statistical test of the proposed scoring system.

The class of correlation models that will be considered here is analogous to the particular class of association models I mentioned earlier. In this class of correlation models, I shall include the following models: (1) a model in which the

correlation between the row classification and column classification is nil (the 0 correlation model); (2) a model with specified (e.g., uniform or unvarying) row-spacing and column-spacing (the U correlation model); (3) a model in which row-spacing is estimated (the R correlation model); (4) a model in which column-spacing is estimated (the C correlation model); and (5) a model in which both row-spacing and column-spacing are estimated (the RC correlation model). As with the association models, Table 4A can be used now to describe the kind of correlation model suited to each kind of cross-classification table.

The correlation models are related to, but different from, the corresponding association models. The statistical methods appropriate for the application of the correlation models are also related to, but different from, the corresponding methods appropriate for the application of the association models. (For example, the test of whether a particular correlation model is congruent with the observed data is related to, but different from, the test of whether the corresponding association model is congruent with the data.) In order to test whether a particular scoring system (e.g., the natural numbers or some other specified scoring system) is suitable for the row categories and/or column categories, we recommend (a) the test obtained using the association models (and the corresponding statistical methods appropriate for them) if the proposed scoring system is supposed to be relevant to the association (as measured in Section 2.1); and (b) the test obtained using the correlation models (and the corresponding statistical methods appropriate for them) if the proposed scoring system is supposed to be relevant to the correlation (as measured in Section 2.2). Some comparisons between the association models and the correlation models will be made in Section 2.3.

The association models can be used to fit a cross-classification table when the row and column classifications arise from underlying continuous random variables having a bivariate normal distribution (if the classification categories are not too wide), and they can be used also under more general circumstances (Goodman, 1981a, 1981b). The correlation models will not be as suitable for this purpose, as we shall see later.

The results obtained with the models considered here are related to results obtained with the canonical correlation approach and the correspondence analysis approach to the analysis of the $I \times J$ cross-classification table. Interest in these approaches has been increasing in recent years (see, e.g., Gabriel, 1971; Benzécri, 1973; Hill, 1974; O'Neill, 1978; Nishisato, 1980; Goodman, 1981a; Haberman, 1981; Escoufier, 1982; Deville and Malinvaud, 1983; Greenacre, 1984; Lebart, Morineau, and Warwick, 1984); and the results I shall present here can be viewed as a contribution to the development of these approaches. (With respect to the evaluation of correspondence analysis, various authors have commented on the need for the further development of the inferential aspects of this approach (see, e.g., Nishisato, 1980, page 205; Aitkin, 1982); and the models and methods that will be introduced in the present lecture should help to meet this need.) The models and methods introduced here can be used to simplify and supplement the analyses based upon these other approaches. To illustrate the utility of the models and methods introduced here, I shall apply them to reanalyze Tables 1 and 2, which were analyzed earlier using the canonical correlation approach and

the correspondence analysis approach, respectively (see Kendall and Stuart, 1979, page 619; and Escoufier, 1982).

The models considered here in Section 2 pertain to the study of association and correlation in the $I \times J$ table. In the special case where $I = J$ and there is a one-to-one correspondence between the row categories and the column categories (see, e.g., Table 3), other aspects of the cross-classification table (e.g., the possible lack of homogeneity between the row classification distribution and the column classification distribution) will often be of interest. To analyze the various possible kinds of asymmetry that may be present in the table, I shall introduce in Section 3 a class of asymmetry models (or generalized symmetry models). This class of models is related to, but different from, the class of generalized independence models I introduced years ago (Goodman 1972). (The class of asymmetry models includes in it some of the models introduced in Goodman, 1979b, 1981c.) The development of the generalized symmetry and generalized independence models then leads us to a class of models that can be used to study simultaneously both asymmetry and nonindependence; viz., the generalized symmetry + independence models. (The class of generalized symmetry + independence models also includes in it some of the models considered in, e.g., Goodman, 1979b, 1981c; see also Hope, 1982; and Sobel, Hout, and Duncan, 1985.) To illustrate the utility of these models, we shall apply them to Table 3.

With the general perspective of this lecture in mind, I shall discuss briefly in Section 4 how to analyze models of the kind considered in the lecture both in the case where the order of a set of related parameters in the model is unspecified and in the case where this order is specified. This topic is touched upon in Section 2 and is developed further in Section 4.

2. Models for row scores and column scores in the $I \times J$ table.

2.1. *Association models.* For the $I \times J$ cross-classification table, let P_{ij} denote the probability that an observation will fall in the ith row and jth column of the table ($i = 1, 2, \cdots, I$; $j = 1, 2, \cdots, J$). The usual model of statistical independence between the row classification and the column classification can be expressed as follows:

$$P_{ij} = \alpha_i \beta_j, \tag{2.1}$$

where α_i and β_j are nonnegative parameters. The α_i and β_j in (2.1) pertain to the row marginals and the column marginals in the $I \times J$ table. (When no entries in the $I \times J$ table are deleted—i.e., when (2.1) applies to all cells (i, j) in the table—the α_i is proportional to the probability that an observation will fall in the ith row, and the β_j is proportional to the probability that an observation will fall in the jth column; but when some specified entries in the table are deleted—i.e., when (2.1) does not apply to all cells (i, j) in the table—the interpretation of the α_i and β_j is somewhat less straightforward. For further details, see, e.g., Goodman, 1968.)

Model (2.1) will be called the null association model (the 0 model). Next we consider some generalizations.

Let μ_i and ν_j denote additional parameters pertaining to the ith row and the

jth column of the $I \times J$ table; and let the probability P_{ij} now be

$$P_{ij} = \alpha_i \beta_j e^{\phi \mu_i \nu_j}. \tag{2.2}$$

The parameter ϕ in (2.2) is a measure of the association in the $I \times J$ table, with $\phi = 0$ for the null association model (the 0 model). The μ_i and ν_j in (2.2) are row scores and column scores, respectively, which will be estimated from the data (i.e., from the observed frequencies in the $I \times J$ table) in the case where the order and spacing of the row categories and/or the column categories is not specified; and these scores will be determined by the specified spacing (and by condition (2.14) below) in the case where the spacing of the row categories and the column categories is specified. (To unify the exposition of the class of association models to be considered here, we have found it convenient to refer to the μ_i and ν_j as parameters, both in the case where the spacing is not specified and in the case where it is.) The meaning of the μ_i, ν_j, and ϕ in (2.2) is clarified further by (2.4), (2.6), and (2.14) below, and by additional comments in my earlier work (Goodman 1979a, 1981a, 1981b).

Model (2.2) is called the RC association model. The class of association models that will be considered here includes the RC association model and various modifications of this model. These association models are related, in one way or another, to models in Birch (1965), Rasch (1966), Haberman (1974a; 1979, page 371), Simon (1974), Duncan (1979), Goodman (1979c), Andersen (1980, page 210), Plackett (1981, page 76), and Clogg (1982a, 1982b). When the association models that will be considered here (including model (2.2)) are reexpressed in terms of the natural logarithm of P_{ij}, we also find that they are somewhat similar to nonadditivity models in the analysis of a two-way array with continuous data, of the kind considered in, e.g., Mandel (1971), Johnson and Graybill (1972), and Tukey (1977, page 421). For some discussion of this similarity, see, e.g., Goodman (1981a). (Although my attention in the present lecture is focused on the contingency table context, it may be worth noting that the development of methods appropriate for the analysis of association models in the contingency table context (as presented here) can also lead to the development of some related methods appropriate for the analysis of nonadditivity models in a two-way array with continuous data, and these related methods will make a new contribution to the statistical tools now available for the analysis of nonadditivity with the analysis of variance; see Goodman and Haberman, 1984.)

I shall now proceed with the development of the association models for the $I \times J$ cross-classification table. Let $\Theta_{ij,i'j'}$ denote the odds-ratio for the 2×2 subtable obtained from the ith and i'th rows and the jth and j'th columns ($i \neq i'$, $j \neq j'$),

$$\Theta_{ij,i'j'} = (P_{ij}P_{i'j'})/(P_{i'j}P_{ij'}); \tag{2.3}$$

and let $\Phi_{ij,i'j'}$ denote the natural logarithm of $\Theta_{ij,i'j'}$. We see from (2.2) and (2.3) that

$$\Phi_{ij,i'j'} = \phi(\mu_i - \mu_{i'})(\nu_j - \nu_{j'}). \tag{2.4}$$

From (2.4), we see that ϕ measures the association $\Phi_{ij,i'j'}$ obtained from the 2×2 subtable when the rows (i and i') are one unit apart ($\mu_i - \mu_{i'} = 1$) and the

columns (j and j') are also one unit apart ($\nu_j - \nu_{j'} = 1$). The parameter ϕ is a measure of "intrinsic association," and it is expressed in units proportional to the scales used to specify μ_i and ν_j.

When the order of the rows and the order of the columns is specified, we rearrange the rows and the columns in the specified order, with $i = 1, 2, \cdots, I$ and $j = 1, 2, \cdots, J$ denoting now the specified order. In this case, we let Θ_{ij} be defined as follows:

$$\Theta_{ij} = \Theta_{ij,i'j'}, \quad \text{for} \quad i' = i + 1, \quad j' = j + 1 \quad (i = 1, 2, \cdots, I - 1; j = 1, 2, \cdots, J - 1); \tag{2.5}$$

and we let Φ_{ij} denote the natural logarithm of Θ_{ij}. (In Anscombe, 1981, page 327, the Θ_{ij} are called Goodman ratios.) Thus, from (2.4) and (2.5) we see that

$$\Phi_{ij} = \phi(\mu_i - \mu_{i+1})(\nu_j - \nu_{j+1}). \tag{2.6}$$

Consider now the RC model (2.2) with, say, the following conditions imposed upon the μ_i and ν_j:

$$\mu_i - \mu_{i+1} = \Delta', \quad \text{and} \quad \nu_j - \nu_{j+1} = \Delta'', \tag{2.7}$$

with the spacing parameters Δ' and Δ'' either specified or unspecified. In (2.7), the rows are equally spaced and the columns are equally spaced. From (2.6) and (2.7), we see that

$$\Phi_{ij} = \phi\Delta'\Delta'' \quad (\text{for } i = 1, 2, \cdots, I - 1; j = 1, 2, \cdots, J - 1). \tag{2.8}$$

Thus, the Φ_{ij} are constant (i.e., uniform). The RC model (2.2) with condition (2.7) imposed is called the U model, and we can also use this name to denote the generalized form of this model obtained when condition (2.7) is replaced by the more general condition:

$$\mu_i - \mu_{i+1} = \Delta_i', \quad \text{and} \quad \nu_j - \nu_{j+1} = \Delta_j'', \tag{2.9}$$

with the spacing parameters Δ_i' and Δ_j'' specified except for possible scale factors (see Goodman, 1979a). In the U model considered here, each difference $\mu_i - \mu_{i+1}$ is specified (and unvarying or fixed) in accordance with (2.9) (except for a possible scale factor), and similarly for each difference $\nu_j - \nu_{j+1}$; with (2.7) being a special case of (2.9).

In a similar way, we obtain the R association model when the following condition (2.10) or the more general condition (2.11) is imposed:

$$\nu_j - \nu_{j+1} = \Delta'', \tag{2.10}$$

or

$$\nu_j - \nu_{j+1} = \Delta_j'', \tag{2.11}$$

with the spacing parameter Δ_j'' specified except for a possible scale factor. And we obtain the C association model when the following condition (2.12) or the more general condition (2.13) is imposed:

$$\mu_i - \mu_{i+1} = \Delta', \tag{2.12}$$

or

$$\mu_i - \mu_{i+1} = \Delta_i', \tag{2.13}$$

with the spacing parameter Δ_i' specified except for a possible scale factor.

Since the general form of the RC model (2.2) remains unchanged when the parameters μ_i and ν_j are replaced by linear functions of the μ_i and ν_j, respectively, the following restrictions can be imposed upon these parameters without any loss of generality:

$$\begin{aligned} &\textstyle\sum_{i=1}^{I} \mu_i P_{i\cdot} = 0, \quad \sum_{j=1}^{J} \nu_j P_{\cdot j} = 0, \\ &\textstyle\sum_{i=1}^{I} \mu_i^2 P_{i\cdot} = 1, \quad \sum_{j=1}^{J} \nu_j^2 P_{\cdot j} = 1, \end{aligned} \tag{2.14}$$

where

$$P_{i\cdot} = \textstyle\sum_{j=1}^{J} P_{ij}, \quad \text{and} \quad P_{\cdot j} = \sum_{i=1}^{I} P_{ij}.$$

These particular restrictions are imposed here for expository purposes, to facilitate comparison later with the results obtained using the correlation models, the canonical correlation approach, and the correspondence analysis approach.

Without loss of generality, any two independent restrictions could have been imposed on the μ_i (e.g., any restriction on the mean of the μ_i (as defined in (2.14)) and any restriction on the second moment of the μ_i (as defined in (2.14))); similarly, any two independent restrictions could also have been imposed on the ν_j. (For an example of restrictions on the μ_i and ν_j that are different from, but related to, the restrictions in (2.14), see, e.g., Goodman, 1979a). Any restrictions of this kind simply replace the μ_i and ν_i by linear functions of the μ_i and ν_j, respectively. The only arbitrariness in the μ_i and ν_j is with respect to their location and scale, and the kind of restrictions considered here simply remove this arbitrariness. The interpretation of the μ_i and ν_j is not seriously affected by this arbitrariness, since it is easy enough to change the μ_i and ν_j to any linear functions of the μ_i and ν_j, respectively. The restrictions in (2.14) simply facilitate comparison with the correlation models, the canonical correlation approach, and the correspondence analysis approach; and they also have some other advantages in certain contexts. (For example, in the situation where the $I \times J$ table is modified by combining a given subset of the rows and/or by combining a given subset of the columns, the μ_i and ν_j in the modified table, with restriction (2.14) imposed, are directly comparable to the corresponding quantities in the original table, with restriction (2.14) imposed, under conditions described in Goodman, 1981a).

By the way, it may also be worth noting here that even in the case where the spacing of rows and columns is specified, some of the observed data (viz., the observed row marginals and column marginals) are used to estimate the μ_i and ν_j when (2.14) is imposed, since these data are relevant in the calculation of the location and scale of the μ_i and ν_j; but when (2.14) is replaced by restrictions of the kind presented in Goodman (1979a), the observed data are not used for this purpose. Since some of the observed data are used here to estimate the μ_i and ν_j even in the case where the spacing of rows and columns is specified, this is an

additional reason why we refer to the μ_i and ν_j as parameters in this case, as well as in the case where the spacing is not specified. (When (2.14) is replaced by restrictions of the kind presented in Goodman (1979a), we need not refer to the μ_i and ν_j as parameters in the case where the spacing is specified, but we still need to refer to them as parameters in the case where the spacing is not specified.)

To illustrate the application of the association models, we present in Table 5 results obtained when these models are applied to Table 1. The goodness-of-fit and likelihood-ratio chi-squared values are given in Table 5A for each of the models, and the maximum-likelihood estimates of the parameters appear in Table 5B. (The U, R, and C models in Table 5 were applied here using (2.7), (2.10), and (2.12), respectively, and the RC model was obtained with (2.2). Methods for calculating maximum-likelihood estimates for the association and correlation models considered in this lecture will be discussed in the Appendix.) The C and RC models fit the data well; and the corresponding estimated scores (viz., the $\hat{\mu}_i$ and $\hat{\nu}_j$) provide an adequate picture of the pattern of association between row and column classifications, using (2.4) to describe this picture with the various quantities in (2.4) replaced by their corresponding maximum-like-

Table 5A
Association models applied to Table 1

association models	degrees of freedom	goodness-of-fit chi-squared	likelihood-ratio chi-squared
0	9	44.34	46.89
U	8	11.13	11.86
R	6	9.29	9.88
C	6	4.29	4.35
RC	4	1.76	1.74

TABLE 5B
Estimated scores, $\hat{\mu}_i$ and $\hat{\nu}_j$, obtained with association models applied to Table 1, the estimated intrinsic association $\hat{\phi}$, and the observed correlation between the $\hat{\mu}_i$ and $\hat{\nu}_j$

	association models			
	U	**R**	**C**	**RC**
intrinsic association	.58	.59	.65	.66
correlation	.49	.51	.54	.56
row $\hat{\mu}$				
1	−1.53	−1.38	−1.53	−1.34
2	−.62	−1.01	−.62	−1.06
3	.30	.49	.30	.46
4	1.21	1.10	1.21	1.14
column $\hat{\nu}$				
1	.96	.96	.83	.85
2	.09	.09	.51	.44
3	−.78	−.78	−1.53	−1.62
4	−1.65	−1.65	−1.19	−1.08

lihood estimates. Since both the C and RC models fit the data well, in comparing these two models we note that the C model is more parsimonious than the RC model, and the corresponding picture of the association pattern obtained using (2.4) is also more parsimonious for the C model.

As we noted earlier, the C model can be used when the order of the row categories is specified and the spacing between them is also specified (see (2.12)–(2.13)), with the order of the column categories unspecified (see Table 4A). We shall now show how to modify this model when the order of the column categories is also specified (but column spacing is unspecified). When the order of the column categories is specified as indicated in, say, Table 1, the $\hat{\nu}_j$ for the C model in Table 5B indicate that the order of the $\hat{\nu}_j$ is not monotonic (see $\hat{\nu}_3$ and $\hat{\nu}_4$). On the other hand, when there is this specified column order, we could impose, say, the following restriction on the ν_j parameters in the C model:

$$\nu_j \geq \nu_{j+1} \quad (\text{for } j = 1, 2, \cdots, J-1). \tag{2.15}$$

Since the likelihood function under the C model is concave (Haberman, 1973), if the unrestricted C model yields estimates $\hat{\nu}_j$ that contradict the restriction—e.g., if $\hat{\nu}_3 < \hat{\nu}_4$ (as in Table 5B)—then the likelihood function will be maximized on the boundary, with $\nu_3 = \nu_4$. When this kind of equality restriction is imposed, the analysis can be carried out simply by applying the C model to the original cross-classification modified by combining columns 3 and 4. Thus, we obtain for the original cross-classification table with restriction (2.15) imposed, $\hat{\nu}_1 = .83$, $\hat{\nu}_2 = .51$, and $\hat{\nu}_3 = \hat{\nu}_4 = -1.36$ under the C model, and $\hat{\nu}_1 = .86$, $\hat{\nu}_2 = .45$, and $\hat{\nu}_3 = \hat{\nu}_4 = -1.35$ under the RC model. (For the sake of completeness, we present in Table 6 results obtained when the various association models are applied to the original Table 1 modified by combining columns 3 and 4.)

If all the $\hat{\nu}_j$ were replaced by $-\hat{\nu}_j$ (i.e., if the $\hat{\nu}_j$ were multiplied by minus one), then the observed correlation would change its sign and there would be no other change in our results. A similar remark applies if all the $\hat{\mu}_i$ were replaced by $-\hat{\mu}_i$. To facilitate comparison with the canonical correlation approach, we have adopted the convention here of using the $\hat{\mu}_i$ if $\hat{\mu}_1 < 0$, and the $-\hat{\mu}_i$ if $\hat{\mu}_1 > 0$. (If $\hat{\mu}_1 = 0$, then we use the $\hat{\mu}_i$ if $\hat{\mu}_2 < 0$, and the $-\hat{\mu}_i$ if $\hat{\mu}_2 > 0$; etc.) We then use either the $\hat{\nu}_j$ or the $-\hat{\nu}_j$, whichever makes the observed correlation nonnegative. The corresponding convention is also adopted in Section 2.2 with our correlation models. On the other hand, to facilitate comparison with the correspondence analysis in Section 2.4, we use there the $\hat{\mu}_i$ if $\hat{\mu}_1 > 0$, and the $-\hat{\mu}_i$ if $\hat{\mu}_1 < 0$. (Compare Tables 5B, 6B, 7B, and 8B with Table 12.)

The above comments pertain to the case where only one of the pairs ($\hat{\nu}_j$ and $\hat{\nu}_{j+1}$, for $j = 1, 2, \cdots, J-1$) contradicted restriction (2.15); viz., the pair $\hat{\nu}_3$ and $\hat{\nu}_4$. These comments can be directly extended to the case where more than one of the pairs contradict the restriction. For related material, see the discussion later in Section 4.

2.2. *Correlation models.* We begin this section by returning for a moment to the usual model of statistical independence between the row classification A and

TABLE 6A
Association models applied to modified Table 1

association models	degrees of freedom	goodness-of-fit chi-squared	likelihood-ratio chi-squared
0′	6	42.86	45.39
U′	5	7.87	7.70
R′	3	4.33	4.51
C′	4	3.11	3.29
RC′	2	1.20	1.18

TABLE 6B
Estimated scores, $\hat{\mu}_i$ and $\hat{\nu}_j$, obtained with association models applied to modified Table 1, *the estimated intrinsic association $\hat{\phi}$, and the observed correlation between the $\hat{\mu}_i$ and $\hat{\nu}_j$*

	association models			
	U′	R′	C′	RC′
intrinsic association	.61	.62	.64	.64
correlation	.51	.53	.54	.55
row $\hat{\mu}$				
1	−1.53	−1.37	−1.53	−1.37
2	−.62	−1.07	−.62	−1.01
3	.30	.59	.30	.45
4	1.21	1.01	1.21	1.14
column $\hat{\nu}$				
1	1.05	1.05	.83	.86
2	−.09	−.09	.51	.45
3	−1.23	−1.23	−1.36	−1.35

the column classification B. This model is usually expressed as follows:

$$P_{ij} = P_{i.}P_{.j}, \tag{2.16}$$

where

$$P_{i.} = \sum_{j=1}^{J} P_{ij}, \quad \text{and} \quad P_{.j} = \sum_{i=1}^{I} P_{ij}.$$

The $P_{i.}$ and $P_{.j}$ are the usual ith row marginal probability and jth column marginal probability, respectively, in the $I \times J$ table. To facilitate the exposition in this section on correlation models, model (2.16) will now be called the null correlation model (the 0 model).

The null correlation model (2.16) is equivalent to the null association model (2.1) in the case where no entries in the $I \times J$ table are deleted (i.e., where (2.1) applies to all cells (i, j) in the table). We shall comment later on the analysis of cross-classification tables in which specified entries in the table are deleted. For expository purposes, we use the parameters α_i and β_j to describe the null association model (2.1), and the parameters $P_{i.}$ and $P_{.j}$ to describe the null correlation model (2.16), since the association models considered here can be easily extended to the analysis of cross-classification tables in which specified entries in the table are deleted (and the formulation in terms of the parameters α_i and β_j is appropriate for this purpose), whereas a corresponding extension for

the class of correlation models is less straightforward (even if we had formulated the null correlation model in more general terms).

We shall next consider generalizations of the null correlation model (2.16). Let x_i and y_j denote additional parameters pertaining to the ith row and jth column of the $I \times J$ table, and let λ denote an additional parameter pertaining to the correlation between x_i and y_j. Using these additional parameters, we now define our first generalization of model (2.16), viz., the RC correlation model

$$P_{ij} = P_{i.}P_{.j}(1 + \lambda x_i y_j), \tag{2.17}$$

where

$$\begin{aligned} &\textstyle\sum_{i=1}^{I} x_i P_{i.} = 0, \quad \sum_{j=1}^{J} y_j P_{.j} = 0, \\ &\textstyle\sum_{i=1}^{I} x_i^2 P_{i.} = 1, \quad \sum_{j=1}^{J} y_j^2 P_{.j} = 1. \end{aligned} \tag{2.18}$$

The parameter λ in (2.17) is a measure of the correlation between x_i and y_j, since

$$\textstyle\sum_{i=1}^{I} \sum_{j=1}^{J} x_i y_j P_{ij} = \lambda, \tag{2.19}$$

under (2.17)–(2.18). When $\lambda = 0$ in (2.17), we obtain the null correlation model (2.16). The x_i and y_j in (2.17) are row scores and column scores, respectively, which will be estimated from the data (i.e., from the observed frequencies in the $I \times J$ table) in the case where the order and spacing of the row categories and/or the column categories is not specified; and these scores will be determined by the specified spacing (and by condition (2.18)) in the case where the spacing of the row categories and the column categories is specified. Together with (2.18)–(2.19), the meaning of λ, x_i, and y_j in (2.17) is clarified further by (2.20)–(2.23) below. (We followed the usual convention of using Greek letters for parameters in our introduction of the parameters μ_i and ν_j in the association model (2.2), but have used English letters for corresponding parameters x_i and y_j in the correlation model (2.17) because this will facilitate comparison with results presented in the earlier literature on canonical correlation and on correspondence analysis. A similar remark could also be made with respect to our use of the English letter P for the probability parameter rather than the Greek Π.)

The x_i and y_j in the RC model (2.17) can yield a picture of the pattern of nonindependence in the $I \times J$ table, since

$$(P_{ij} - P_{i.}P_{.j})/(P_{i.}P_{.j}) = \lambda x_i y_j, \tag{2.20}$$

and

$$(P_{ij} - P_{i.}P_{.j})^2/(P_{i.}P_{.j}) = \lambda^2 x_i^2 y_j^2 P_{i.}P_{.j}. \tag{2.21}$$

From (2.18) and (2.21), we obtain

$$\textstyle\sum_{i=1}^{I} \sum_{j=1}^{J} (P_{ij} - P_{i.}P_{.j})^2/(P_{i.}P_{.j}) = \lambda^2(\sum_{i=1}^{I} x_i^2 P_{i.})(\sum_{j=1}^{J} y_j^2 P_{.j}) = \lambda^2, \tag{2.22}$$

with the partitioning of the goodness-of-fit measure (2.22) into components pertaining to the relative contribution $x_i^2 P_{i.}$ of the ith row (for $i = 1, 2, \cdots, I$) or to the relative contribution $y_j^2 P_{.j}$ of the jth column (for $j = 1, 2, \cdots, J$).

From (2.19) we saw that the parameter λ in the RC correlation model (2.17) is the correlation between the row scores x_i and column scores y_j; and now we

see from (2.22) that the square of λ is equal to the goodness-of-fit measure (2.22) under model (2.17). In addition, we see that the row scores x_i and column scores y_j also pertain to the partitioning of the goodness-of-fit measure. The total goodness-of-fit measure (2.22) can be partitioned into components pertaining to the ith row (viz., $x_i^2 P_{i.}\lambda^2$) for $i = 1, 2, \cdots, I$; and it can also be partitioned into components pertaining to the jth column (viz., $y_j^2 P_{.j}\lambda^2$), for $j = 1, 2, \cdots, J$. From (2.21) we see that the product of the ith row component and the jth column component (viz., $x_i^2 y_j^2 P_{i.} P_{.j}\lambda^4$) is proportional to the component pertaining to the cell (i, j) in the $I \times J$ table, for $i = 1, 2, \cdots, I$ and $j = 1, 2, \cdots J$. Additional insight into the meaning of the x_i and y_j is also obtained from (2.23) below.

From (2.17)–(2.18), we see that

$$\sum_{j=1}^{J} y_j(P_{ij}/P_{i.}) = \lambda x_i, \tag{2.23a}$$

and

$$\sum_{i=1}^{I} x_i(P_{ij}/P_{.j}) = \lambda y_j. \tag{2.23b}$$

Thus, the row scores x_i are proportional to the weighted average of the y_j, using the conditional distribution in the ith row as weights; and the column scores y_j are proportional to the weighted average of the x_i, using the conditional distribution in the jth column as weights.

As earlier with association models in Section 2.1, we now introduce the conditions:

$$x_i - x_{i+1} = \Delta_i^*, \quad \text{and} \quad y_j - y_{j+1} = \Delta_j^{**}, \tag{2.24}$$

or just

$$y_j - y_{j+1} = \Delta_j^{**}, \tag{2.25}$$

or just

$$x_i - x_{i+1} = \Delta_i^*, \tag{2.26}$$

with the spacing parameters Δ_i^* and Δ_j^{**} specified except for a possible scale factor. (Compare (2.24)–(2.26) with (2.9), (2.11), and (2.13).) The U model, the R model, and the C model are obtained from the RC model (2.17), by imposing conditions (2.24), (2.25), and (2.26), respectively.

Let f_{ij} denote the observed frequency in the ith row and jth column of the $I \times J$ table, and let N denote the sample size; $N = \sum_{i=1}^{I} \sum_{j=1}^{J} f_{ij}$. Let p_{ij} denote the corresponding observed proportion f_{ij}/N, with $p_{i.}$ and $p_{.j}$ defined by

$$p_{i.} = \sum_{j=1}^{J} p_{ij}, \quad \text{and} \quad p_{.j} = \sum_{i=1}^{I} p_{ij}.$$

Let $\hat{P}_{ij}$ denote the maximum-likelihood estimate of P_{ij} under a particular model, with $\hat{P}_{i.}$ and $\hat{P}_{.j}$ defined by

$$\hat{P}_{i.} = \sum_{j=1}^{J} \hat{P}_{ij}, \quad \text{and} \quad \hat{P}_{.j} = \sum_{i=1}^{I} \hat{P}_{ij}.$$

Under the RC correlation model (2.17), we find that

$$\hat{P}_{i.} = p_{i.}, \quad \text{and} \quad \hat{P}_{.j} = p_{.j}; \tag{2.27}$$

but these equalities will generally not hold true for the U correlation model (with (2.24) satisfied). With the R correlation model, the second equality in (2.27) will hold true, but the first equality will generally not hold true; and with the C model, the first equality in (2.27) will hold true, but the second equality will generally not hold true. Because of this, I also introduce here estimated "approximate correlation models," by obtaining maximum-likelihood estimates of parameters in the models subject to the restriction (2.27), for the U, R, and C models. For further details, see the Appendix.

To illustrate the application of the correlation models, we present in Table 7 results obtained when these models are applied to Table 1. (The U, R, and C models in Table 7 were applied here using (2.24), (2.25), and (2.26) respectively, with $\Delta_i^* = \Delta^*$ and $\Delta_j^{**} = \Delta^{**}$.) The results obtained with the approximate correlation models in Table 7 are generally similar to those obtained with the corresponding correlation models, and they would be even more similar in cases where $|\lambda|$ is smaller. On the other hand, estimates obtained with the approximate correlation model are not efficient estimates under the corresponding correlation model (when $\lambda \neq 0$), and so the values in parentheses in Table 7A, which were

TABLE 7A

Correlation models and approximate correlation models applied to Table* 1

correlation models	degrees of freedom	goodness-of-fit chi-squared	likelihood-ratio chi-squared
0	9	44.34	46.89
U	8	13.20 (13.31)	14.37 (14.42)
R	6	9.59 (9.65)	10.71 (10.74)
C	6	5.46 (5.52)	5.91 (5.95)
RC	4	1.85	1.83

* Values obtained with approximate correlation models are in parentheses.

TABLE 7B

Estimated scores, $\hat{x}_i$ and $\hat{y}_j$, obtained with correlation models applied to Table 1, *and the estimated correlation $\hat{\lambda}$ between the $\hat{x}_i$ and $\hat{y}_j$*

	correlation models				canonical score
	U	R	C	RC	
correlation	.43	.49	.50	.56	.56
row $\hat{x}$					
1	−1.54	−1.34	−1.53	−1.36	−1.39
2	−.63	−1.16	−.62	−1.10	−1.06
3	.29	.59	.30	.61	.60
4	1.20	.99	1.21	.99	1.00
column $\hat{y}$					
1	.97	.95	.82	.84	.84
2	.10	.09	.58	.49	.48
3	−.77	−.78	−1.45	−1.55	−1.58
4	−1.64	−1.65	−1.24	−1.17	−1.14

obtained with the approximate correlation models, will not have asymptotic chi-squared distributions.

As we noted earlier with the association models in Table 5, we now find with the correlation models that the C and RC models fit the data well; and the corresponding estimated scores, $\hat{x}_i$ and $\hat{y}_j$, provide an adequate picture of the pattern of nonindependence between the row and column classifications, using (2.20)–(2.22) to describe this picture, with the various quantities in (2.20)–(2.22) replaced by their corresponding maximum-likelihood estimates. Since both C and RC correlation models fit the data well, we note (as we did earlier with the association models) that the C model is more parsimonious; and, as we noted earlier, in the case where there is specified column order with unspecified column spacing (in addition to the specified row order and specified row spacing in the C model), the nonmonotonicity of the $\hat{y}_j$ in Table 6B under the C model (see $\hat{y}_3$ and $\hat{y}_4$) leads us to an analysis of the original cross-classification modified by combining columns 3 and 4. This analysis is included here in Table 8.

Some of the results obtained here in our analysis of Table 1 and the modified Table 1 can be summarized in a table that partitions the total chi-squared value (i.e., the chi-squared value under the 0 model) into appropriate components. This was done for association models in Goodman (1979a, 1981a). Now we show how

TABLE 8A

Correlation models and approximate correlation models applied to modified Table* 1

correlation models	degrees of freedom	goodness-of-fit chi-squared	likelihood-ratio chi-squared
0′	6	42.86	45.39
U′	5	9.07 (9.32)	9.78 (9.92)
R′	3	4.95 (5.10)	5.40 (5.53)
C′	4	4.47 (4.51)	4.82 (4.86)
RC′	2	1.18	1.16

* Values obtained for approximate correlation models are in parentheses.

TABLE 8B

Estimated scores, $\hat{x}_i$ and $\hat{y}_j$, obtained with correlation models applied to modified Table 1, *and the estimated correlation $\hat{\lambda}$ between the $\hat{x}_i$ and $\hat{y}_j$*

	correlation models				canonical score
	U′	R′	C′	RC′	
correlation	.47	.52	.50	.56	.56
row $\hat{x}$					
1	−1.50	−1.24	−1.53	−1.36	−1.41
2	−.59	−1.16	−.62	−1.09	−1.02
3	.32	.57	.30	.59	.59
4	1.23	1.08	1.21	1.02	1.01
column $\hat{y}$					
1	1.04	1.05	.82	.84	.84
2	−.10	−.09	.58	.49	.49
3	−1.24	−1.23	−1.34	−1.36	−1.36

to combine the partitioning obtained for Table 1 and for the modified Table 1 (which we shall now call Table 1′ or T_1') into a single partitioning for the association models (Table 9A) and a corresponding single partitioning for the correlation models (Table 9B). The numerical results in Table 9A were obtained using entries from Tables 5A and 6A, and the corresponding results in Table 9B were obtained using entries from Table 7A and Table 8A. Tables 9A and 9B focus our attention on the analysis of the components of association and correlation in Table 1′ (i.e., in T_1') and on the component for heterogeneity between columns 3 and 4 in Table 1 (i.e., in T_1). These partitionings shed light on the possible homogeneity between columns 3 and 4 in Table 1, on the adequacy of the RC models [(2.2) and (2.17)] in Table 1′, on the possible equal spacing of the μ_i and the x_i, and on the intrinsic association and correlation in Table 1′ using unequally spaced ν_j and y_j and equally spaced μ_i and x_i.

With respect to the model for the homogeneity between columns 3 and 4 in Table 1, this model can be tested by using the difference between the likelihood-ratio chi-squared values obtained for the 0 model in Table 1 and the 0′ model in Table 1′, as with the fourth component in Tables 9A and 9B. (The 0′ model refers here to the 0 model applied to Table 1′. For the fourth component in Tables 9A and 9B, the difference between the chi-squared values for models 0 and 0′ is referred to, in an abbreviated notation, in the second column of the tables.) An equivalent (but more direct) test of the homogeneity model could

TABLE 9A

Analysis of association (ANOAS) in Table 1 and Table 1′: components of association

components	models used	degrees of freedom	likelihood-ratio chi-squared
General effect and column effect in T_1'	0′ − C′	6 − 4 = 2	42.10
Row effect in T_1'	C′ − RC′	4 − 2 = 2	2.11
Other effects in T_1'	RC′	2	1.18
Heterogeneity between cols. 3 and 4 in T_1	0 − 0′	9 − 6 = 3	1.50
Total effects in T_1	0	9	46.89

TABLE 9B

Analysis of correlation (ANOCOR) in Table 1 and Table 1′: components of correlation

components	models used	degrees of freedom	likelihood-ratio chi-squared
General effect and column effect in T_1'	0′ − C′	6 − 4 = 2	40.57
Row effect in T_1'	C′ − RC′	4 − 2 = 2	3.66
Other effects in T_1'	RC′	2	1.16
Heterogeneity between cols. 3 and 4 in T_1	0 − 0′	9 − 6 = 3	1.50
Total effects in T_1	0	9	46.89

have been obtained by using the chi-squared value for a test of independence in the subtable formed from Table 1 by deleting all cells that are not included in columns 3 and 4. (For further development of the corresponding two general approaches to the analysis of homogeneity, see, e.g., Goodman 1968, 1981d.) With respect to the other components in Tables 9A and 9B, they are all obtained by direct application of the general partitioning approach to the analysis of association models in Goodman (1979a, 1981a).

Somewhat different partitionings are presented next in Tables 10A and 10B. These tables shed light on the adequacy of the RC models in Table 1, on the possible equal spacing of the μ_i and x_i, on the possible equality $\nu_3 = \nu_4$ and $y_3 = y_4$, and on the intrinsic association and correlation in Table 1 using unequally spaced ν_j (with $\nu_3 = \nu_4$) and y_j (with $y_3 = y_4$) and equally spaced μ_i and x_i.

Some comments about the interpretation of the components in Table 10 may be helpful. The C model in Table 1 with $\nu_3 = \nu_4$ is equivalent to the combined model of homogeneity between columns 3 and 4 and the C′ model in Table 1′ (see Goodman 1981a, 1981d). A test pertaining to the general effect and column effect in the C model in Table 1 with $\nu_3 = \nu_4$ (i.e., a test of null association in Table 1, assuming that the C model with $\nu_3 = \nu_4$ holds true) is obtained by comparing the chi-squared value for the 0 model in Table 1 with the chi-squared

TABLE 10A
Analysis of association (ANOAS) in Table 1 and Table 1′: components of association

components	models used	degrees of freedom	likelihood-ratio chi-squared
General effect and column effect in C model with $\nu_3 = \nu_4$ in T_1	0′ − C′	6 − 4 = 2	42.10
Inequality between ν_3 and ν_4 in C model in T_1	0 − 0′ + C′ − C	7 − 6 = 1	0.44
Row effect in RC model in T_1	C − RC	6 − 4 = 2	2.61
Other effects in RC model in T_1	RC	4	1.74
Total effects in T_1	0	9	46.89

TABLE 10B
Analysis of correlation (ANOCOR) in Table 1 and Table 1′: components of correlation

components	models used	degrees of freedom	likelihood-ratio chi-squared
General effect and column effect in C model with $y_3 = y_4$ in T_1	0′ − C′	6 − 4 = 2	40.57
Inequality between y_3 and y_4 in C model in T_1	0 − 0′ + C′ − C	7 − 6 = 1	0.41
Row effect in RC model in T_1	C − RC	6 − 4 = 2	4.08
Other effects in RC model in T_1	RC	4	1.83
Total effects in T_1	0	9	46.89

value for the combined model of homogeneity between columns 3 and 4 (viz., the difference between the chi-squared values obtained for models 0 and 0′) and the C′ model. The test obtained is therefore based on the chi-squared statistic

$$X^2(0) - \{[X^2(0) - X^2(0')] + X^2(C')\} = X^2(0') - X^2(C'), \tag{2.28}$$

where $X^2(M)$ denotes the chi-squared value obtained under model M. Similarly, a test pertaining to the component for the possible inequality between ν_3 and ν_4 in the C model in Table 1 (i.e., a test of the null hypothesis that $\nu_3 = \nu_4$ in the C model) is obtained by comparing the chi-squared value for the combined model described above with the chi-squared value for the C model; thus obtaining

$$\{[X^2(0) - X^2(0')] + X^2(C')\} - X^2(C). \tag{2.29}$$

(Formulae (2.28) and (2.29) are referred to, in abbreviated notation, in the second column of Tables 10A and 10B, for the first and second components, respectively.) Note that the sum of the statistics (2.28) and (2.29) is

$$\begin{aligned}[X^2(0') - X^2(C')] + [X^2(0) - X^2(0') + X^2(C') - X^2(C)] \\ = X^2(0) - X^2(C).\end{aligned} \tag{2.30}$$

The statistics (2.28) and (2.29) can serve as the components in a partitioning of the statistic (2.30), where (2.30) is used as a test pertaining to the general effect and column effect in the usual C model in Table 1.

We noted in Goodman (1981a, 1981d) that formulae of the kind described by (2.28) and (2.29) can be extended in a straightforward way to study the possible equality of a given subset of the parameters ν_j. In this case, these formulae can be applied directly, using as the modified cross-classification table (say, a Table 1′) the original table modified by combining the columns pertaining to the given subset. Also, with respect to the parameters ν_j in the C model, if the columns of the original table are partitioned into mutually exclusive and exhaustive subsets, in a study of the possible equality of the ν_j that are included within each subset, the modified cross-classification table would be the original table modified by combining the columns included within each subset. (With respect to Table 1, in the example I first used here, the particular application of (2.28)–(2.29) that was presented could have been described as a partitioning of the four columns into three mutually exclusive subsets, where the first two subsets consist of one column each and the third subset consists of two columns (columns 3 and 4).) Similarly with respect to the parameters ν_j and μ_i in the RC model, if the columns of the original table are partitioned into mutually exclusive and exhaustive subsets, and the rows are also partitioned into mutually exclusive and exhaustive subsets, in a study of the possible equality of the ν_j that are included within each subset of the columns and the possible equality of the μ_i that are included within each subset of the rows, formulae (2.29) can be applied directly, replacing C and C′ in this formula by RC and RC′, respectively; and using as the modified cross-classification table the original table modified by combining the columns that are included within each subset of the columns and also combining the rows included within each subset of the rows.

In closing this discussion of the extension of formulae (2.28) and (2.29), we return for a moment to our earlier comment on the equivalence between the statistic for testing the homogeneity model (viz., $X^2(0) - X^2(0')$), which was included in these formulae, and the corresponding statistic obtained with a more direct test of the homogeneity model. If the columns and the rows of the original table are partitioned as described above, then the homogeneity model in this context states that the columns included within each subset of columns are homogeneous, and the rows included within each subset of rows are homogeneous. With the explicit formulae for the maximum-likelihood estimate of the expected frequencies under this homogeneity model (see, e.g., Goodman, 1981d), the corresponding chi-squared statistic can be calculated directly (comparing the observed frequencies in the original cross-classification table with these estimated expected frequencies), and this statistic is equivalent to the statistic $X^2(0) - X^2(0')$. The chi-squared statistic calculated directly reduces, in certain special cases, to a statistic for testing independence or quasi-independence applied to a subtable formed from the original cross-classification table by deleting the cells that are not relevant to the particular homogeneity model. (For further details, see Goodman 1981a, 1981d.) In some special cases, it will be simpler to apply the chi-squared statistic calculated directly (or the equivalent statistic for testing independence or quasi-independence applied to the appropriate subtable of the original $I \times J$ table); and in other cases, we would calculate the equivalent difference statistic, $X^2(0) - X^2(0')$.

The number of degrees of freedom for testing the homogeneity model is simply the difference $(I-1)(J-1) - (I'-1)(J'-1)$, where I' and J' are the number of rows and the number of columns, in the modified table (see, e.g., Goodman, 1981d). (When $I' = I$, this difference is equal to $(I-1)(J-J')$.) The corresponding number of degrees of freedom for the statistic (2.28) is the difference

$$(I'-I)(J'-1) - (I'-2)(J'-1) = J'-1;$$

and similarly the number of degrees of freedom for the statistic (2.30) is $J-1$. In view of the relationship between (2.28), (2.29), and (2.30), we see that the number of degrees of freedom for the statistic (2.29) is simply $J-J'$. When the C and C′ in formula (2.29) are replaced by RC and RC′, respectively, we obtain $I+J-I'-J'$ degrees of freedom for the statistic (2.29). (Note also that the number of degrees of freedom for the term in brackets in formulae (2.29) is equal to

$$(J-J') + (I-2)(J-1) = (I-1)(J-1) - (J'-1) = IJ - I - J - J' + 2;$$

and similarly the corresponding number of degrees of freedom is equal to

$$\begin{aligned}(I+J-I'-J') + (I-2)(J-2) &= (I-1)(J-1) - (I'+J'-3)\\ &= IJ - I - J - I' - J' + 4,\end{aligned}$$

when the C′ in the term in brackets in (2.29) is replaced by RC′.)

2.3. *Comparison between association models and correlation models.*

2.3.1. *Introductory comments.* Let us begin this section with a comparison between the association models and correlation models in the simple case of the 2×2 table. In this case, the U, R, C, RC models are all saturated models, and estimation is thus simplified. Also, for the 2×2 table, we see from (2.14) and (2.18) that

$$\begin{aligned} \mu_1 &= x_1 = -(P_{2.}/P_{1.})^{1/2}, \quad \mu_2 = x_2 = (P_{1.}/P_{2.})^{1/2}, \\ \nu_1 &= y_1 = -(P_{.2}/P_{.1})^{1/2}, \quad \nu_2 = y_2 = (P_{.1}/P_{.2})^{1/2}. \end{aligned} \tag{2.31}$$

Under the association models, we see from (2.6) and (2.31) that

$$\hat{\phi} = \{\log[(p_{11}p_{22})/(p_{12}p_{21})]\}(p_{1.}p_{2.}p_{.1}p_{.2})^{1/2}; \tag{2.32}$$

and under the correlation models, we see from (2.19) (or from (2.20) and (2.31)) that

$$\hat{\lambda} = (p_{11}p_{22} - p_{12}p_{21})/(p_{1.}p_{2.}p_{.1}p_{.2})^{1/2}. \tag{2.33}$$

As we have already noted in the more general $I \times J$ table, the $\hat{\phi}$ pertains to the logarithm of the odds-ratios (see (2.6)), while $\hat{\lambda}$ pertains to the correlation (see (2.19)).

When $\hat{\lambda}$ is close to zero, the estimates (2.32) and (2.33) will be close to each other; but they can differ greatly when $\hat{\lambda}$ is not close to zero. In the more general $I \times J$ table, we see from (2.2) and (2.17) that the association models and the corresponding correlation models will give similar results when λ (or ϕ) is close to zero; but the results can be quite different when λ is not close to zero. (Of course, when $\lambda = \phi = 0$, models (2.2) and (2.17) are identical.)

In my first article comparing the association models and the canonical correlation approach (Goodman, 1981a), I described conditions under which the $\hat{\mu}_i$ and $\hat{\nu}_j$ of the association model (2.2) (with $\phi \neq 0$) will be approximately equal to the $\hat{x}_i$ and $\hat{y}_j$ of the correlation model (2.17). These conditions can be characterized as follows: (a) when a sample cross-classification table arises from an underlying (discretized) joint bivariate normal; or (b) when it arises from an underlying bivariate distribution that is not necessarily normal, but where separate transformations of the row scores and the column scores can be found so that the transformed scores have a discretized joint bivariate normal distribution (see Goodman, 1981a).

It is interesting to note that the association model (2.2) (with $\phi \neq 0$) can be quite different from the correlation model (2.17), even when the $\hat{\mu}_i$ and $\hat{\nu}_j$ of model (2.2) are approximately equal to the $\hat{x}_i$ and $\hat{y}_j$ of model (2.17)—e.g., even when the conditions described in the preceding paragraph are satisfied. To illustrate this point, we first note that in order to describe a bivariate normal distribution, the tetrachoric series based upon Tchebycheff–Hermite polynomials can be used (see, Lancaster, 1957, 1969; and Kendall and Stuart, 1979, page 602), and application of only the first term in the series will usually not be adequate

(unless the correlation is close to zero). Since the correlation model (2.17) corresponds only to the first term in the tetrachoric series, this model will usually not be adequate as a description of the cross-classification arising from an underlying bivariate normal. On the other hand, the association model (2.2) can provide an excellent approximation to the discretized bivariate normal, as we shall see in Section 2.3.2. Thus we find that when the cross-classification arises from, say, an underlying discretized bivariate normal (with row and column scores pertaining to the corresponding levels of the underlying variables), the estimated scores $\hat{x}_i$ and $\hat{y}_j$ obtained from model (2.17), and the corresponding estimated scores $\hat{\mu}_i$ and $\hat{\nu}_j$ obtained from model (2.2), will be approximately equal to the corresponding row and column scores (see Goodman, 1981a); but the bivariate distribution obtained with the correlation model (2.17) will not provide as good an approximation of the distribution in the cross-classification table as will the association model (2.2).

There is a direct relationship between the parameter ϕ in (2.2) and the correlation between the row scores and column scores, μ_i and ν_j, when the cross-classification table arises from an underlying bivariate normal, and also under some more general conditions (see Goodman, 1981a). Even when these conditions are not satisfied, the parameter ϕ in (2.2) will be relevant as a measure of intrinsic association (see (2.4)), but the correlation between the row scores and column scores will not be particularly meaningful under (2.2).

In comparing the association model (2.2) with the correlation model (2.17), we note that the association model (2.2) is defined for all possible values of μ_i and ν_j; whereas the correlation model (2.17) is defined only when

$$x_i y_j \geq -1/|\lambda|, \quad \text{for} \quad i = 1, 2, \cdots, I; \quad j = 1, 2, \cdots, J. \tag{2.34}$$

Condition (2.34) can be very limiting, as will be evident if an attempt is made to apply the correlation model in, say, the examples of Section 2.3.2 (see Table 11), where condition (2.34) is violated. In addition, as noted earlier, with the identifying condition (2.14) applied in the association model, there is no loss of generality; but this is not the case, in a certain sense, when condition (2.18) is applied in the correlation model. (If the right-hand side of (2.17) is replaced by $\alpha_i\beta_j(1 + \lambda x_i y_j)$, the corresponding conditions pertaining to zero mean in (2.18) introduce a loss of generality.)

Here is another way in which association models are preferable to correlation models: The methods developed for maximum-likelihood estimation and testing using association models can be directly extended to the case where the $I \times J$ table has deleted entries (which are missing, void, unreliable, or restricted in certain ways); see, e.g., Goodman (1979a). But the correlation models are defined in a way that would make it more difficult to extend these models to the case where there are deleted entries in the $I \times J$ table, without changing the character of the models to some extent. (With the simplest correlation model (where $\lambda = 0$), if there are deleted entries in the $I \times J$ table, the formulation of the correlation model would be changed to the same formulation used with the corresponding association model in order to make it applicable; compare (2.16)–(2.17) with (2.1)–(2.2).) It is possible to generalize the correlation models in order

to make an extended version of them applicable to the case where there are deleted entries in the $I \times J$ table, but this is less straightforward than the corresponding generalization of the association model, and it is also beyond the scope of this lecture. In addition, although we noted earlier that the association model (2.2) and the corresponding correlation model (2.17) will give similar results when λ (or ϕ) is close to zero, the results can be quite different if any of the $x_i y_j$ is sufficiently negative (and large in absolute value) so that condition (2.34) is either violated or almost violated. (In this lecture, when we comment on the case where λ is close to zero, we shall limit consideration to the case where $\lambda x_i y_j$ is also sufficiently close to zero, for $i = 1, 2, \cdots, I$; $j = 1, 2, \cdots, J$.)

Before closing this section, we return briefly to (2.32) and (2.33). These formulae for maximum-likelihood estimates of ϕ and λ, respectively, were expressed explicitly in terms of the observed proportions p_{ij}. There was no need to use iterative procedures in the calculation of these maximum-likelihood estimates, since the association models and correlation models (U, R, C, and RC) are all saturated models for the 2×2 table (see Table 4B). Similarly, since the R and RC models are saturated models for the $I \times 2$ table, and the C and RC models are saturated models for the $2 \times J$ table, explicit formulae that are generalizations of (2.32) and (2.33) can be obtained in these cases. For example, from (2.22) we obtain

$$\hat{\lambda} = [\textstyle\sum_{i=1}^{I} \sum_{j=1}^{J} (p_{ij} - p_{i\cdot}p_{\cdot j})^2/(p_{i\cdot}p_{\cdot j})]^{1/2}, \tag{2.35}$$

in the cases noted above (viz., the $I \times 2$ table and the $2 \times J$ table, with the corresponding saturated models). The sign of $\hat{\lambda}$ in (2.35) is arbitrary in a certain sense (since the sign of λ in (2.19) changes when the x_i are replaced by $-x_i$, or when the y_j are replaced by $-y_j$), but we take here as the sign of $\hat{\lambda}$ the sign obtained with (2.33) applied to a 2×2 subtable (formed from, say, rows 1 and 2 and columns 1 and 2) of the full $I \times J$ table. If the sign obtained with (2.33) is different for different 2×2 subtables, then we can take whichever sign predominates when considering all subtables formed from, say, rows i and $i + 1$ and columns j and $j + 1$, with $i = 1, 2, \cdots, I - 1$ and $j = 1, 2, \cdots, J - 1$. (The sign of $\hat{\lambda}$ obtained in this way can differ from the sign obtained following the usual convention in canonical correlation analysis; but in the present context, the method used here is preferable to some extent. With respect to the usual convention in canonical correlation analysis, where either the $\hat{y}_j$ or the $-\hat{y}_j$ are used, whichever makes the correlation $\hat{\lambda}$ nonnegative, we could instead decide between the $\hat{y}_j$ and the $-\hat{y}_j$ following the same general method used in deciding between the $\hat{x}_i$ and the $-\hat{x}_i$, and the sign of the correlation $\hat{\lambda}$ would be a direct consequence of this decision.)

For the $I \times J$ table with $I = 2$ and/or $J = 2$, further simplification of (2.35) will shed some additional light on the relationship between the $\hat{\lambda}$ $\hat{x}_i$, and $\hat{y}_j$ obtained with the RC correlation model. We shall illustrate this point for the case where $I = 2$ and $J \geq 2$. In this case, the $\hat{x}_i$ are given by (2.31); and from (2.23b) we see that $\hat{y}_j$ can be expressed as

$$\hat{y}_j = \hat{y}_j'/\hat{\lambda}, \tag{2.36}$$

with

$$\hat{y}_j' = (p_{1\cdot}p_{2j} - p_{2\cdot}p_{1j})/[p_{\cdot j}(p_{1\cdot}p_{2\cdot})^{1/2}], \tag{2.37}$$

and

$$\hat{\lambda} = [\textstyle\sum_{j=1}^{J} (\hat{y}_j')^2 p_{\cdot j}]^{1/2}. \tag{2.38}$$

The $\hat{y}_j'$ can also be written as

$$\begin{aligned} \hat{y}_j' &= -(p_{1j} - p_{1\cdot}p_{\cdot j})/[p_{\cdot j}(p_{1\cdot}p_{2\cdot})^{1/2}] \\ &= (p_{2j} - p_{2\cdot}p_{\cdot j})/[p_{\cdot j}(p_{1\cdot}p_{2\cdot})^{1/2}]; \end{aligned} \tag{2.39}$$

see, e.g., (2.20).

Similarly, with respect to the relationship between $\hat{\phi}$, $\hat{\mu}_i$, and $\hat{\nu}_j$ for the RC association model in this case (with $I = 2$), the $\hat{\mu}_i$ are given by (2.31); and from (2.6) we see that $\hat{\nu}_j$ can be expressed as

$$\hat{\nu}_j = \hat{\nu}_j'/\hat{\phi}, \tag{2.40}$$

with

$$\hat{\nu}_{j+1}' - \hat{\nu}_j' = \hat{\Phi}_{1j}(p_{1\cdot}p_{2\cdot})^{1/2}, \tag{2.41}$$

and

$$\hat{\phi} = [\textstyle\sum_{j=1}^{J} (\hat{\nu}_j')^2 p_{\cdot j}]^{1/2}. \tag{2.42}$$

The $\hat{\nu}_j'$ can be written as $\hat{\nu}_j' = \hat{\nu}_j''(p_{1\cdot}p_{2\cdot})^{1/2}$, with

$$\hat{\nu}_j'' = \begin{cases} \sum_{k=1}^{J-1} \hat{\Phi}_{1k}(1 - \sum_{t=1}^{k} p_{\cdot t}), & \text{for } j = 1 \\ \hat{\nu}_1'' - \sum_{k=1}^{j-1} \hat{\Phi}_{1k}, & \text{for } j > 1; \end{cases} \tag{2.43a}$$

or equivalently as

$$\hat{\nu}_j'' = \begin{cases} \sum_{k=j}^{J-1} \hat{\Phi}_{1k} + \hat{\nu}_J'', & \text{for } j = 1, 2, \cdots, J-1 \\ -\sum_{k=1}^{J-1} \hat{\Phi}_{1k}(\sum_{t=1}^{k} p_{\cdot t}), & \text{for } j = J. \end{cases} \tag{2.43b}$$

These formulae can be expressed also in other equivalent terms; e.g.,

$$\hat{\nu}_1'' = \textstyle\sum_{k=1}^{J-1} \Phi_{1k}(\sum_{t=k+1}^{J} p_{\cdot t}), \quad \hat{\nu}_J'' = -\sum_{k=1}^{J-1} \hat{\Phi}_{1k}(1 - \sum_{t=k+1}^{J} p_{\cdot t}),$$

etc.

From the results in the preceding two paragraphs, we find, among other things, that the difference between the estimated ν_j is a function of the log-odds $\hat{\Phi}_{1j}$; i.e.,

$$\hat{\nu}_{j+1} - \hat{\nu}_j = \hat{\Phi}_{1j}(p_{1\cdot}p_{2\cdot})^{1/2}/\hat{\phi}; \tag{2.44}$$

and that the estimate of y_j is a function of the difference between observed and expected frequencies (under the null hypothesis of independence), weighted inversely by $p_{\cdot j}$; i.e.,

$$\hat{y}_j = (p_{1\cdot}p_{2j} - p_{2\cdot}p_{1j})/[p_{\cdot j}(p_{1\cdot}p_{2\cdot})^{1/2}\hat{\lambda}]. \tag{2.45}$$

2.3.2. *Association models and a generalized bivariate normal with arbitrary marginals.* Let us begin this section by considering first the usual bivariate normal distribution in which the univariate marginal distributions are, of course, also normally distributed. Various authors have compared this bivariate normal with Plackett's model (see, e.g., Plackett, 1965; Mardia, 1967; and Anscombe, 1981, page 307), and have concluded that the two distributions are close to one another. A particular example of Plackett's model was compared with the bivariate normal distribution having correlation $\rho = 0.5$ in Pearson (1913) and Plackett (1965), $\rho = 0.45$ in Mardia (1967), and $\rho = 0.6$ in Anscombe (1981, page 307). With $\rho = 0.5$, the central ordinate of each cumulative distribution function is equated; with $\rho = 0.6$, the central ordinate of each density function is equated; and with $\rho = 0.45$, a special kind of correlation coefficient calculated for each distribution is equated (see Mardia, 1967). In the section on approximation to the bivariate normal distribution in Mardia (1967, Section 7), he concludes that his method (viz., equating a special kind of correlation coefficient calculated for each distribution) is "better than Plackett's method [viz., equating the central ordinate of each cumulative distribution] everywhere except at points in the neighborhood of (0, 0). However, ... if the D.F. [distribution function] around (0, 0) is of importance then Plackett's method is recommended." (In addition, using Plackett's distribution as an approximation to the bivariate normal, Mardia presents methods for estimating the correlation coefficient of an underlying bivariate normal distribution in the $I \times J$ table; see Mardia, 1967, Section 9; 1970, Section 8.7.) Anscombe (1981, page 306) first compares the bivariate normal and Plackett's model by equating the central ordinate of each density function; and he then suggests that, for the particular example of Plackett's model under consideration, a value of ρ in the neighborhood of 0.4 be chosen. (He equates moments to obtain this value.)

For the bivariate normal distribution with $\rho = 0.5$, comparison of Plackett's model with the association models shows that the latter models agree more closely with the normal than does Plackett's model (Goodman, 1981b). Now Tables 11A, 11B, and 11C yield similar results for $\rho = 0.6$, $\rho = 0.45$, and $\rho = 0.4$. In addition, with respect to the estimation of the correlation coefficient, we find that in all these cases, the estimate $\hat{\phi}$, when transformed as described in Goodman (1981b), yields estimates of ρ that agree with the true ρ to two decimal places; and a similar result is obtained with an estimate of ρ based upon the correlation between the row scores and column scores, $\hat{\mu}_i$ and $\hat{\nu}_j$, estimated under the model.

Plackett's model did not agree so well with the bivariate normal because of the skewness of the conditional distributions obtained with this model and the nonlinearity of its regressions (Pearson, 1913), and because "quadrant association" is constant in Plackett's model but not in the bivariate normal (Mosteller, 1968). Of course, if we were to compare Plackett's model and the association models with respect to how well they approximate some other bivariate distribution (different from the bivariate normal), the results obtained would depend upon the particular distribution under consideration.

Let X and Y now denote continuous random variables, and let $\mu(X)$ and $\nu(Y)$ denote functions of X and Y that have a standard bivariate normal distribution

TABLE 11A

Bivariate normal distribution with $\rho = 0.6$ (first entry) compared with the corresponding uniform association model (second entry) and Plackett's model (third entry). Probability = Entry/10^4

y	x					
	0.0	**0.5**	**1.0**	**1.5**	**2.0**	**2.5**
	3	8	13	15	12	10
	3.7	8.7	13.9	15.2	11.5	7.7
	11	12	14	7	5	1
2.5						
	17	32	40	35	21	12
	16.8	31.4	40.3	35.4	21.4	11.5
	27	36	33	23	8	5
2.0						
	68	101	104	73	35	15
	67.7	101.3	104.0	73.1	35.4	15.2
	76	97	88	52	24	6
1.5						
	188	224	184	104	40	13
	187.7	224.7	184.7	104.0	40.3	13.9
	180	210	165	88	32	15
1.0						
	356	340	224	101	32	8
	356.5	341.7	224.7	101.3	31.4	8.7
	340	331	210	97	36	12
0.5						
	463	356	188	68	17	3
	464.7	356.5	187.7	67.7	16.8	3.7
	458	340	180	77	26	11
0.0						
	415	256	109	32	6	1
	415.7	255.3	107.6	31.1	6.2	1.1
	397	239	115	49	17	6
−0.5						
	256	127	43	10	2	0
	255.3	125.5	42.3	9.8	1.6	0.2
	238	131	64	26	10	3
−1.0						
	109	43	12	2	0	0
	107.6	42.3	11.4	2.1	0.3	0.0
	116	63	30	14	4	2
−1.5						
	32	10	2	0	0	0
	31.1	9.8	2.1	0.3	0.0	0.0
	49	27	13	5	2	1
−2.0						
	6	2	0	0	0	0
	6.2	1.6	0.3	0.0	0.0	0.0
	17	9	5	2	1	0
−2.5						
	1	0	0	0	0	0
	1.1	0.2	0.0	0.0	0.0	0.0
	6	3	2	0	1	0

TABLE 11B

Bivariate normal distribution with $\rho = 0.45$ (first entry) compared with the corresponding uniform association model (second entry) and Plackett's model (third entry). Probability = Entry/10^4

y	x					
	0.0	**0.5**	**1.0**	**1.5**	**2.0**	**2.5**
	7	12	14	12	8	5
	7.7	12.0	13.8	11.7	7.3	4.3
	11	12	14	7	5	1
2.5						
	26	35	35	26	14	8
	25.9	35.3	35.4	26.1	14.3	7.3
	27	36	33	23	8	5
2.0						
	82	97	84	54	26	12
	81.7	97.1	84.8	54.6	26.1	11.7
	76	97	88	52	24	6
1.5						
	191	197	150	84	35	14
	190.9	197.9	151.0	84.8	35.4	13.8
	180	210	165	88	32	15
1.0						
	328	296	197	97	35	12
	328.8	297.5	197.9	97.1	35.3	12.0
	340	331	210	97	36	12
0.5						
	416	328	191	82	26	7
	416.6	328.8	190.9	81.7	25.9	7.7
	458	340	180	77	26	11
0.0						
	389	269	137	51	14	3
	389.2	268.0	135.8	50.7	14.0	3.6
	397	239	115	49	17	6
−0.5						
	269	162	72	24	6	1
	268.0	161.1	71.2	23.2	5.6	1.3
	238	131	64	26	10	3
−1.0						
	137	72	28	8	2	0
	135.8	71.2	27.5	7.8	1.6	0.3
	116	63	30	14	4	2
−1.5						
	51	24	8	2	0	0
	50.7	23.2	7.8	1.9	0.4	0.1
	49	27	13	5	2	1
−2.0						
	14	6	2	0	0	0
	14.0	5.6	1.6	0.4	0.1	0.0
	17	9	5	2	1	0
−2.5						
	3	1	0	0	0	0
	3.6	1.3	0.3	0.1	0.0	0.0
	6	3	2	0	1	0

TABLE 11C

Bivariate normal distribution with $\rho = 0.4$ (first entry) compared with the corresponding uniform association model (second entry) and Plackett's model (third entry). Probability = Entry/10^4

y	x					
	0.0	**0.5**	**1.0**	**1.5**	**2.0**	**2.5**
	9	12	13	11	6	4
	9.0	12.6	13.2	10.3	6.0	3.4
	11	12	14	7	5	1
2.5						
	28	35	33	23	12	6
	28.4	35.3	33.0	23.1	12.1	6.0
	27	36	33	23	8	5
2.0						
	85	94	79	49	23	11
	85.0	94.5	78.9	49.3	23.1	10.3
	76	97	88	52	24	6
1.5						
	191	190	141	79	33	13
	190.9	189.6	141.5	78.9	33.0	13.2
	180	210	165	88	32	15
1.0						
	321	285	190	94	35	12
	320.7	284.5	189.6	94.5	35.3	12.6
	340	331	210	97	36	12
0.5						
	405	321	191	85	28	9
	404.7	320.7	190.9	85.0	28.4	9.0
	458	340	180	77	26	11
0.0						
	383	271	144	57	17	4
	382.5	270.8	144.0	57.3	17.1	4.8
	397	239	115	49	17	6
−0.5						
	271	171	81	29	8	2
	270.8	171.2	81.3	28.9	7.7	2.0
	238	131	64	26	10	3
−1.0						
	144	81	34	11	3	1
	144.0	81.3	34.5	11.0	2.6	0.6
	116	63	30	14	4	2
−1.5						
	57	29	11	3	1	0
	57.3	28.9	11.0	3.1	0.7	0.1
	49	27	13	5	2	1
−2.0						
	17	8	3	1	0	0
	17.1	7.7	2.6	0.7	0.1	0.0
	17	9	5	2	1	0
−2.5						
	4	2	1	0	0	0
	4.8	2.0	0.6	0.1	0.0	0.0
	6	3	2	0	1	0

with correlation coefficient ρ. Then the probability density function for (X, Y) can be expressed as

$$f(x, y) = \alpha(x, \rho)\beta(y, \rho)e^{\phi'\mu(x)\nu(y)}, \tag{2.46}$$

where

$$\phi' = \rho/(1 - \rho^2), \tag{2.47}$$

and

$$\begin{aligned} \alpha(x, \rho) &= \exp(-\mu^2(x)/2(1 - \rho^2))(d\mu/dx)(2\pi)^{-1/2}(1 - \rho^2)^{-1/4} \\ \beta(y, \rho) &= \exp(-\nu^2(y)/2(1 - \rho^2))(d\nu/dy)(2\pi)^{-1/2}(1 - \rho^2)^{-1/4}, \end{aligned} \tag{2.48}$$

with $d\mu/dx$ and $d\nu/dy$ the derivatives of $\mu(x)$ and $\nu(y)$. (Here we consider functions $\mu(x)$ and $\nu(y)$ that are differentiable and have inverses.) Consider the class of all random variables (X, Y) that have density functions of the general form (2.46)–(2.48), i.e., all random variables (X, Y) that can be expressed in terms of separate transformations of random variables that have a joint bivariate normal distribution with correlation coefficient ρ. The corresponding density functions form the translation family obtained from the bivariate normal (see, Mardia, 1970, page 30). This family of density functions can be further extended by considering

$$f(x, y) = \alpha^*(x, \rho)\beta^*(y, \rho)e^{\phi'\mu(x)\nu(y)}, \tag{2.49}$$

where $\alpha^*(x, \rho)$ and $\beta^*(y, \rho)$ are nonnegative functions of (x, ρ) and (y, ρ), respectively, which need not necessarily satisfy (2.48). The only restriction on $\alpha^*(x, \rho)$ and $\beta^*(y, \rho)$ is that the resulting $f(x, y)$ be a density function, with the integral over the range of (x, y) equal to one. This family of density functions is richer than the translation family, and will prove more useful in various ways. The association models can be viewed as the discrete analogue of this family of density functions.

To further clarify the difference between the family of density functions (2.46)–(2.48) (i.e., the translation family) and the more general family (2.49) introduced above, we note that the restrictions (2.48) for the translation family can also be expressed as follows:

$$\begin{aligned} \alpha(x, \rho) &= \alpha'(x)\exp(-\mu^2(x)\rho^2/2(1 - \rho^2))(1 - \rho^2)^{-1/4} \\ \beta(y, \rho) &= \beta'(y)\exp(-\nu^2(y)\rho^2/2(1 - \rho^2))(1 - \rho^2)^{-1/4}, \end{aligned} \tag{2.50}$$

with

$$\begin{aligned} \alpha'(x) &= \exp(-\mu^2(x)/2)(d\mu/dx)(2\pi)^{-1/2} \\ \beta'(y) &= \exp(-\nu^2(x)/2)(d\nu/dy)(2\pi)^{-1/2}, \end{aligned} \tag{2.51}$$

where $\alpha'(x)$ and $\beta'(y)$ are the marginal density functions for the random variables X and Y, respectively. With the more general family (2.49), the $\alpha^*(x, \rho)$ and $\beta^*(y, \rho)$ need not be of the same form as $\alpha(x, \rho)$ and $\beta(y, \rho)$ in (2.50)–(2.51).

2.3.3. *Correlation models, association models, and generalized bivariate multinomial distributions.* With the usual multinomial distribution applied to the

$I \times J$ table, the marginal distributions for the row classification and for the column classification will be multinomial distributions, with $P_{i.}$ being the probability that an observation will fall in the ith row category ($i = 1, 2, \cdots, I$), and $P_{.j}$ being the probability that an observation will fall in the jth column category ($j = 1, 2, \cdots, J$). The correlation model (2.17) yields a bivariate multinomial distribution, with the corresponding marginal multinomial distributions defined in terms of the $P_{i.}$ and $P_{.j}$ in (2.17). This bivariate distribution is a direct generalization of the bivariate binomial distribution considered by Aitken and Gonin (1935) and Hamdan and Martinson (1971). Formula (2.35) for the $I \times J$ table with $I = 2$ or $J = 2$ is a generalization of a corresponding result obtained by Hamdan and Martinson (1971) for the 2×2 table.

The association model (2.2) yields a different bivariate multinomial, with the corresponding marginal distributions defined in terms of

$$P_{i.} = \alpha_i \sum_{j=1}^{J} \beta_j e^{\phi \mu_i \nu_j}, \quad P_{.j} = \beta_j \sum_{i=1}^{I} \alpha_i e^{\phi \mu_i \nu_j}. \tag{2.52}$$

As we noted in Section 2.3.2, the association model (2.2) can also be viewed as a discretized version of a generalization of the bivariate normal in which the marginal distributions are not necessarily normal.

2.4. *Correlation models, association models, canonical correlation, and correspondence analysis.* We shall show in this section how the association models and correlation models can simplify and supplement results obtained with canonical correlation and correspondence analysis.

For an $I \times J$ cross-classification table, let $M = \min(I - 1, J - 1)$. The canonical correlation model is

$$P_{ij} = P_{i.}P_{.j}(1 + \sum_{m=1}^{M} \lambda_m x_{im} y_{jm}), \tag{2.53}$$

where

$$\begin{aligned} \sum_{i=1}^{I} x_{im} P_{i.} &= 0, & \sum_{j=1}^{J} y_{jm} P_{.j} &= 0, \\ \sum_{i=1}^{I} x_{im}^2 P_{i.} &= 1, & \sum_{j=1}^{J} y_{jm}^2 P_{.j} &= 1, \\ \sum_{i=1}^{I} x_{im} x_{im'} P_{i.} &= 0, & \sum_{j=1}^{J} y_{jm} y_{jm'} P_{.j} &= 0, \end{aligned} \tag{2.54}$$

for $m \neq m'$. The parameter λ_m in (2.53) is a measure of the correlation between x_{im} and y_{jm}, since

$$\sum_{i=1}^{I} \sum_{j=1}^{J} x_{im} y_{jm} P_{ij} = \lambda_m, \tag{2.55}$$

under (2.53)–(2.54). Model (2.53) will be called the saturated RC correlation model. We next introduce the unsaturated RC correlation models

$$P_{ij} = P_{i.}P_{.j}(1 + \sum_{m=1}^{M^*} \lambda_m x_{im} y_{jm}), \tag{2.56}$$

where $1 \leq M^* \leq M$. The case where $M^* = 1$ was considered in (2.17), and the methods used for this case can be directly extended to the more general case. See the Appendix herein; and, for related material, see Williams (1952), Kendall and Stuart (1979, page 599), and Gilula and Haberman (1984).

We can also extend the association models considered earlier. In this case, the

saturated RC association model is

$$P_{ij} = \alpha_i \beta_j \exp\left(\sum_{m=1}^{M} \phi_m \mu_{im} \nu_{jm}\right), \tag{2.57}$$

where the μ_{im} and ν_{jm} in (2.57) satisfy the same kind of conditions as described by (2.54). (Compare (2.57) with (2.53).) Similarly, we can introduce next the unsaturated RC association models

$$P_{ij} = \alpha_i \beta_j \exp\left(\sum_{m=1}^{M^*} \phi_m \mu_{im} \nu_{jm}\right), \tag{2.58}$$

where $1 \leq M^* \leq M$. Here, too, the methods developed earlier for the case where $M^* = 1$ (see (2.2)) can be directly extended. For related material, see Goodman (1981a), and Chuang, Gheva, and Odoroff (1983).

Table 1, analyzed in Sections 2.1 and 2.2, is a 4×4 table, so $M = 3$. The last column of Table 7B gives the $\hat{\lambda}_1$, $\hat{x}_{i1}$, and $\hat{y}_{j1}$ obtained with the saturated correlation model ($M = 3$) for the 4×4 table. With our earlier analysis of the unsaturated RC model with $M^* = 1$ (see Table 7A), we found that it was not necessary to consider $M^* = 2$ or $M^* = 3$ in the 4×4 table. Also, we found with our more detailed analysis of the case where $M^* = 1$, using the U, R, C, and RC models, that further simplification was possible. A similar comment applies also to our earlier analysis of the modified Table 1 (see Table 8A).

Now let us consider Table 2. This table was analyzed earlier using correspondence analysis in Escoufier (1982). Correspondence analysis is a reparameterization of canonical correlation analysis. In this case, we can express the correspondence analysis model as

$$P_{ij} = P_{i\cdot} P_{\cdot j} \left(1 + \sum_{m=1}^{M} x'_{im} y'_{jm} / \lambda_m\right), \tag{2.59}$$

where

$$x'_{im} = \lambda_m x_{im}, \quad y'_{jm} = \lambda_m y_{jm}. \tag{2.60}$$

From (2.54) and (2.60), we see that

$$\begin{aligned} &\sum_{i=1}^{I} x'_{im} P_{i\cdot} = 0, \qquad \sum_{j=1}^{J} y'_{jm} P_{\cdot j} = 0, \\ &\sum_{i=1}^{I} x'^2_{im} P_{i\cdot} = \lambda_m^2, \quad \sum_{j=1}^{J} y'^2_{jm} P_{\cdot j} = \lambda_m^2. \end{aligned} \tag{2.61}$$

This is the saturated model, and we can introduce an unsaturated version as we did earlier in (2.56):

$$P_{ij} = P_{i\cdot} P_{\cdot j} \left(1 + \sum_{m=1}^{M^*} x'_{im} y'_{jm} / \lambda_m\right), \tag{2.62}$$

for $1 \leq M^* \leq M$.

As earlier, we see that the correlation between x'_{im} and y'_{jm} is equal to λ_m,

$$\sum_{i=1}^{I} \sum_{j=1}^{J} x'_{im} y'_{jm} P_{ij} / \lambda_m^2 = \lambda_m, \tag{2.63}$$

and also

$$\begin{aligned} &\sum_{i=1}^{I} \sum_{j=1}^{J} (P_{ij} - P_{i\cdot} P_{\cdot j})^2 / P_{i\cdot} P_{\cdot j} \\ &\qquad = \sum_{m=1}^{M} \left(\sum_{i=1}^{I} x'^2_{im} P_{i\cdot}\right)\left(\sum_{j=1}^{J} y'^2_{jm} P_{\cdot j}\right) / \lambda_m^2 = \sum_{m=1}^{M} \lambda_m^2. \end{aligned} \tag{2.64}$$

Thus, with the saturated model, the goodness-of-fit measure (2.64), which is

used in testing the null correlation model ($\lambda_m = 0$, for $m = 1, 2, \cdots, M$), can be partitioned into M components λ_m^2 (for $m = 1, 2, \cdots, M$); and from (2.61) we see that the component λ_m^2 can be partitioned into I components $x_{im}'^2 P_{i\cdot}$ (for $i = 1, 2, \cdots, I$) or into J components $y_{jm}'^2 P_{\cdot j}$ (for $j = 1, 2, \cdots, J$). Since Table 2 is a 4×6 table, we have $M = 3$. Table 12 gives the correspondence analysis for the saturated model. The proper value (or characteristic value) in this table is λ_m^2 (for $m = 1, 2, 3$). From Table 12 we see that 94% of the goodness-of-fit chi-squared value is due to λ_1^2 in the first dimension (see (2.64)). With our statistical analysis of the unsaturated correlation models (see Table 13), we find that it is sufficient to consider only $M^* = 1$; and, in addition, further simplification is possible. The $\hat{x}_{i1}'$ scores can be equalized for $i = 2, 3$, and the $\hat{y}_{j1}'$ scores can be equalized for $j = 1, 2$ and for $j = 3, 4$, without any great loss. Also, from Table 13 we see that the RC model can be replaced by models C, R, or U, where model U is the most parsimonious, for the data in Table 2 unmodified; and a similar comment also applies to the data in Table 2 modified as indicated in the second section of Table 13.

To facilitate comparison between the correspondence analysis and the correlation models, the row coordinates and column coordinates for the correlation models in Table 12 were given as $\hat{x}_i'$ and $\hat{y}_j'$, where $\hat{x}_i' = \hat{\lambda}\hat{x}_i$ and $\hat{y}_j' = \hat{\lambda}\hat{y}_j$ (see (2.60)). Also, to facilitate comparison, we present in Table 12 the proper value (or characteristic value) $\hat{\lambda}^2$, rather than the correlation $\hat{\lambda}$ (see (2.63)–(2.64)). If we had instead been comparing here the canonical correlation approach with the correlation models considered in this lecture, we would have compared the row and column canonical scores with the $\hat{x}_i$ and $\hat{y}_j$, respectively (see Tables 7B and

TABLE 12

Correspondence analysis and related results obtained with correlation models applied to Table 2

	correspondence analysis			correlation models			
				RC	RC′	C′	U′
dimension	1	2	3				
proper value	.0260	.0014	.0003	.0266	.0257	.0257	.0253
percentage	94%	5%	1%	96%	93%	93%	91%
row coordinates x'							
1	.260	.012	.023	.261	.262	.261	.258
2	.030	.024	−.019	.031	.011	.012	.012
3	−.013	−.069	−.002	−.014	.011	.012	.012
4	−.236	.019	.016	−.241	−.235	−.236	−.234
column coordinates y'							
1	.181	−.018	.028	.177	.185	.185	.179
2	.185	−.011	−.026	.191	.185	.185	.179
3	.059	−.021	−.010	.060	.019	.019	.019
4	−.008	.042	.011	−.009	.019	.019	.019
5	−.164	.044	−.009	−.165	−.167	−.167	−.142
6	−.287	−.061	.005	−.293	−.285	−.286	−.303

* The RC′, C′, and U′ correlation models were applied with restrictions imposed on columns 1 and 2, columns 3 and 4, rows 2 and 3.

TABLE 13
Association models and correlation models applied to Table 2

models	degrees of freedom	association models		correlation models	
		goodness-of-fit chi-squared	likelihood-ratio chi-squared	goodness-of-fit chi-squared	likelihood-ratio chi-squared
for Table 2 unmodified					
0	15	45.99	47.42	45.99	47.42
U	14	9.73	9.89	9.50	9.64
R	12	6.29	6.28	5.83	5.82
C	10	6.78	6.83	5.76	5.78
RC	8	3.57	3.57	2.74	2.75
for Table 2 modified, with cols. 1 and 2 combined, cols. 3 and 4 combined, rows 2 and 3 combined					
0′	6	42.04	43.44	42.04	43.44
U′	5	1.27	1.27	.39	.40
R′	4	1.09	1.09	.39	.40
C′	3	.98	.98	.12	.12
RC′	2	.86	.87	.12	.12
for Table 2 with cols. 1 and 2 interchanged					
0	15	45.99	47.42	45.99	47.42
U	14	9.10	9.21	8.45	8.52
R	12	6.02	6.03	5.10	5.10
C	10	6.78	6.83	5.76	5.78
RC	8	3.57	3.57	2.74	2.75

8B); and we would also have presented the estimated correlation $\hat{\lambda}$ for comparison with the usual principal canonical correlation.

Some of the results obtained here in our analysis of Table 2 and the modified Table 2 (which we shall call Table 2′) can be summarized in Tables 14 and 15. The interpretation of the components in Tables 14 and 15 is somewhat similar in form to the corresponding interpretation presented earlier for components in Tables 9 and 10 (see, e.g., (2.28)–(2.30)).

With respect to the heterogeneity component (i.e., the third component) in Tables 14A and 14B, we see that the homogeneity model for columns 1 and 2, columns 3 and 4, and rows 2 and 3 is congruent with the data. This kind of analysis was also applied to the homogeneity model for columns 5 and 6 (included with the other pairs of columns and rows listed above), and this model, too, was congruent with the data (see Gilula, 1984); but I did not include this model in our analysis here because there is actually enough heterogeneity between columns 5 and 6 in the data so that models that incorporated the supposed homogeneity of columns 5 and 6 turned out to provide, in some respects, a somewhat less adequate description of the data. For example, while the first RC correlation model in Table 12 (i.e., the RC model without homogeneity restrictions) explained, in a certain sense, 96% of the total goodness-of-fit chi-squared value, the corresponding RC model that incorporated the supposed homogeneity of columns 5 and 6 (together with the other pairs of columns and rows listed above)

TABLE 14A
Analysis of association (ANOAS) in Table 2 and Table 2′: components of association

components	models used	degrees of freedom	likelihood-ratio chi-squared
General effects in T_2'	$0' - U'$	$6 - 5 = 1$	42.17
Other effects in T_2'	U'	5	1.27
Heterogeneity between cols. 1 and 2, cols. 3 and 4, rows 2 and 3 in T_2	$0 - 0'$	$15 - 6 = 9$	3.98
Total effects in T_2	0	15	47.42

TABLE 14B
Analysis of correlation (ANOCOR) in Table 2 and Table 2′: components of correlation

components	models used	degrees of freedom	likelihood-ratio chi-squared
General effect in T_2'	$0' - U'$	$6 - 5 = 1$	43.04
Other effects in T_2'	U'	5	0.40
Heterogeneity between cols. 1 and 2, cols. 3 and 4, rows 2 and 3 in T_2	$0 - 0'$	$15 - 6 = 9$	3.98
Total effects in T_2	0	15	47.42

explained only 89%. (While the maximum-likelihood estimate $\hat{\lambda}_1^2$ for the proper value (or characteristic value) λ_1^2 was .0266 for the first RC correlation model in Table 12, it was only .0246 for the corresponding RC model that incorporated the supposed homogeneity of columns 5 and 6, together with the other pairs of columns and rows.) Although the heterogeneity between columns 5 and 6 in the data was not statistically significant, and models that incorporated the supposed homogeneity were more parsimonious, these models are not included here since, in this section on the comparison between correspondence analysis and an analysis based upon the correlation models or association models (described earlier herein), I include models that would be of greater interest to those who are more at home with correspondence analysis.

In the case where the order of the column categories is specified as indicated in Table 2, the $\hat{y}_j$ for the first RC correlation model in Table 12 indicate that the order of the $\hat{y}_j$ is not monotonic (see $\hat{y}_1$ and $\hat{y}_2$), and similar results (not included in Table 12) are obtained with the corresponding C correlation model, and with the RC and C association models. When the order of the column categories is specified, we can impose restriction (2.15) on the column parameters in the association models, or a corresponding restriction on the parameters in the correlation models. As in the earlier analysis of Table 1, the imposition of restriction (2.15) now would lead us to an analysis of Table 2 modified by combining columns 1 and 2. We have omitted this analysis here because it

TABLE 15A
Analysis of association (ANOAS) in Table 2 and Table 2′: components of association

components	models used	degrees of freedom	likelihood-ratio chi-squared
General effect in U model with $\nu_1 = \nu_2$, $\nu_3 = \nu_4$, $\mu_2 = \mu_3$ in T_2	0′ − U′	6 − 5 = 1	42.17
Row and column effects in RC model with $\nu_1 = \nu_2$, $\nu_3 = \nu_4$, $\mu_2 = \mu_3$ in T_2	U′ − RC′	5 − 2 = 3	0.40
Inequality between ν_1 and ν_2, ν_3 and ν_4, μ_2 and μ_3 in RC model in T_2	0 − 0′ + RC′ − RC	11 − 8 = 3	1.28
Other effects in RC model in T_2	RC	8	3.57
Total effects in T_2	0	15	47.42

TABLE 15B
Analysis of correlation (ANOCOR) in Table 2 and Table 2′: components of correlation

components	models used	degrees of freedom	likelihood-ratio chi-squared
General effect in U model with $y_1 = y_2$, $y_3 = y_4$, $x_2 = x_3$ in T_2	0′ − U′	6 − 5 = 1	43.04
Row and column effects in RC model with $y_1 = y_2$, $y_3 = y_4$, $x_2 = x_3$ in T_2	U′ − RC′	5 − 2 = 3	0.28
Inequality between y_1 and y_2, y_3 and y_4, x_2 and x_3 in RC model in T_2	0 − 0′ + RC′ − RC	11 − 8 = 3	1.35
Other effects in RC model in T_2	RC	8	2.75
Total effects in T_2	0	15	47.42

is similar in form to the corresponding analysis carried out earlier herein for Table 1.

In the case where the order of the column categories is not specified (see Table 4A), we can use the estimated parameters in the RC model (or the C model) to suggest a possible order for the column categories. Thus, for example, the estimated column parameters in the RC model in Table 12 would suggest that columns 1 and 2 be interchanged. When this is done, the cross-classification table thus obtained can be analyzed using the general approach presented in this lecture. Some results are included in the final section of Table 13. From these results, we see that in this case the interchange of columns 1 and 2 yields a slight improvement. (Compare the chi-squared values in Table 13 for the U model obtained for Table 2 unmodified and for Table 2 with columns 1 and 2 interchanged.) Although the improvement is slight in this particular example, the

general approach suggested here can sometimes yield dramatic improvements when the order of the categories is not specified.

2.5. *An overview and some additional models.* Each class of models considered here—the association models and the correlation models—includes within it a large number of models, starting first with the most parsimonious (the null 0 model, (2.1) or (2.16)), then with the second most parsimonious (the U model, (2.2) with (2.9), or (2.17) with (2.24)), and proceeding step by step to the least parsimonious (the saturated RC model, (2.57) or (2.53)). To distinguish between the different unsaturated models, we now let $(RC)_1$ denote the unsaturated RC models introduced in Sections 2.1 and 2.2 (viz., (2.2) and (2.7)). Thus, these $(RC)_1$ models are obtained from (2.58) and (2.56), with $M^* = 1$. In a similar way, we can define $(RC)_2$, $(RC)_3$, etc.

The relationships between these models, with respect to their implications and degrees of freedom, are displayed in Figure 1, for the null 0 model and the various unsaturated models with $M^* = 1$ and $M^* = 2$. (For other displays of this general kind, see, e.g., Goodman, 1973, 1978.) Also included in Figure 1 are

Association models and correlation models	Degrees of freedom*
0	(I-1)(J-1)
U	IJ-I-J
R	(I-1)(J-2)
C	(I-2)(J-1)
$(RC)_1$	(I-2)(J-2)
(R+C)	
(U+RC)	IJ-2I-2J+3
(R+RC)	(I-2)(J-3)
(C+RC)	(I-3)(J-2)
$(RC)_2$	(I-3)(J-3)
(R+C+RC)	

* When $I = 2$ and/or $J = 2$, see relevant comments in Section 2.5.

FIG. 1. *The relationship between some models with respect to their implications and degrees of freedom, for the $I \times J$ table.*

additional related models that have not yet been discussed. These will be considered briefly next.

For expository purposes, we shall focus our attention again now on the association models. (The corresponding correlation models could also have been presented here.) We begin with the R + C model in Figure 1, which we shall describe first in terms of the Φ_{ij}, the natural logarithm of the odds-ratio Θ_{ij} in (2.5). The R + C model can be described as

$$\Phi_{ij} = (\psi + \eta_i' + \eta_j'')\Delta_i'\Delta_j'', \tag{2.65}$$

where ψ is a general-effect parameter, η_i' is a row-effect parameter, and η_j'' is a column-effect parameter, and where Δ_i' and Δ_j'' are specified except for possible scale factors. (The R + C model described in Goodman (1981a) is a special case of (2.65) in which the Δ's are all equal to a (specified or unspecified) constant, say, Δ.) Since the effect parameters ψ, η_i', and η_j'' in (2.65) are additive effects, this model will also be called here the R + C model (or, more specifically, the $(R + C)_1$ model). From (2.6), (2.9), (2.10), and (2.11), we see that the U, R, and C models are all special cases of (2.65).

We next consider the following generalization of (2.65):

$$\Phi_{ij} = (\psi + \eta_i' + \eta_j'' + \phi^*\gamma_i'\gamma_j'')\Delta_i'\Delta_j'', \tag{2.66}$$

where the parameters ϕ^*, γ_i', and γ_j'' in (2.66) are multiplicative factors that combine to form an additional term in (2.66). (The R + C + RC model in Goodman (1981a) is a special case of (2.66) in which the Δ's are all equal to a (specified or unspecified) constant, say, Δ.) Model (2.66) will also be called here the R + C + RC model (or, more specifically, the R + C + $(RC)_1$ model). From (2.6) we see that the RC model is a special case of (2.66) (with $\psi = 0$, $\eta_i' = 0$, $\eta_j'' = 0$); and from (2.65) we see, of course, that the R + C model is also a special case of (2.66) (with $\phi^* = 0$). Similarly, we obtain the U + RC model from (2.66) with $\eta_i' = 0$, $\eta_j'' = 0$ (for $i = 1, 2, \cdots, I - 1$; $j = 1, 2, \cdots, J - 1$), the R + RC model from (2.66) with $\eta_j'' = 0$ ($j = 1, 2, \cdots, J - 1$), and the C + RC model from (2.66) with $\eta_i' = 0$ ($i = 1, 2, \cdots, I - 1$).

The U + RC model can now be written as

$$\Phi_{ij} = (\psi + \phi^*\gamma_i'\gamma_j'')\Delta_i'\Delta_j'', \tag{2.67a}$$

or equivalently as

$$\Phi_{ij} = (\bar{\psi} + \xi_i' + \xi_j'' + \bar{\phi}\xi_i'\xi_j'')\Delta_i'\Delta_j'', \tag{2.67b}$$

where

$$\begin{aligned} \bar{\psi} &= \psi + \phi^*, & \bar{\phi} &= 1/\phi^*, \\ \xi_i' &= (\gamma_i' - 1)\phi^*, & \xi_j'' &= (\gamma_j'' - 1)\phi^*. \end{aligned} \tag{2.68}$$

From (2.67a) we see that the RC model is a special case of the U + RC model (with $\psi = 0$); and from (2.67b) we see that the U + RC model can also be viewed as a generalization of the R + C model in which the additional parameter $\bar{\phi}$ has been introduced in (2.67b). Formulae (2.67a)–(2.67b) are a direct extension of

corresponding formulae in Goodman (1981a). In (2.67)–(2.68), the γ_i' and γ_i'' have been normed (without loss of generality) so that $(\sum_i \gamma_i')/I = 1$, $(\sum_j \gamma_j'')/J = 1$, and the ϕ^* has been adjusted accordingly; and a similar remark applies also to the corresponding formulae in Goodman (1981a). (For a related result in the analysis of a two-way array with continuous data, see the discussion of Tukey's test for additivity in, e.g., Tukey, 1977, page 421.)

With respect to the additional models considered in this section (the R + C, U + RC, R + RC, C + RC, and R + C + RC models), we have described above how these models are related to one another; and an overview of these relationships is included in Figure 1. In this overview, we have also included the models discussed in the earlier sections (the 0, U, R, C, $(RC)_1$, and $(RC)_2$ models). Figure 1 could be extended further, in a straightforward way, to include additional models of the general kind we have considered here (e.g., U + $(RC)_2$, R + $(RC)_2$, C + $(RC)_2$, R + C + $(RC)_2$, $(RC)_3$, U + $(RC)_3$, etc.).

Before closing this section, we note that the formulae for the degrees of freedom in Figure 1 should be interpreted with caution. There are actually zero degrees of freedom when a formula in Figure 1 yields either zero or a negative value. (In the special case of the 2×2 cross-classification table ($I = J = 2$), the last formula in Figure 1 is inapplicable.) For further details, see Goodman (1981a).

Models (2.65)–(2.67) were expressed here in terms of the Φ_{ij}; but they could have been expressed equivalently in terms of the P_{ij}. (Compare, e.g., (2.6) with (2.2).) In this case, we would obtain (2.58) with $M^* = 2$ for (2.65) and (2.67), and with $M^* = 3$ for (2.66); and with appropriate conditions placed upon the μ_{im} and ν_{jm} in (2.58). (In the R + C model (2.65) and the R + C + RC model (2.66), the ψ term can be ignored by including it as part of the η_i' or η_j'' terms.) Appropriate formulae expressed in terms of the P_{ij} are direct extensions of the corresponding formulae in Goodman (1981a).

3. Models for generalized symmetry, independence, and symmetry + independence in the $K \times K$ table.

3.1. *Introductory comments.* The general approach considered in Section 2 can be viewed as providing a model building apparatus for the analysis of association and/or correlation in the $I \times J$ table (see, e.g., Tables 4A and 4B, and Figure 1). In the special case where $I = J$, and where there is a one-to-one correspondence between the row categories and column categories in the square cross-classification table, this feature of the table can be taken into account by modifying, in various ways, some of the models considered in Section 2 above. (See, for example, the RC model modified to have homogeneous row-column scores, which was introduced in Goodman, 1979a.) We next present a somewhat different general approach, which will provide a model-building apparatus specifically designed for the analysis of square cross-classification tables of this general kind. We shall be concerned here with the situation where the order of the categories is specified, and shall consider both the case where the order of a set of related parameters is unspecified and the case where it is specified.

3.2. *Generalized symmetry models.* We begin this section by considering first

the usual symmetry model for the square table:

$$P_{ij} = \rho_{ij}, \quad \text{with} \quad \rho_{ij} = \rho_{ji}, \quad \text{for} \quad i < j, \quad i \neq j. \tag{3.1}$$

To facilitate the exposition in this discussion of generalized symmetry models, model (3.1) will now be called the null (0) asymmetry model.

A simple generalization of model (3.1) is then the triangle (T) asymmetry model,

$$P_{ij} = \begin{cases} \rho_{ij}\tau_1, & \text{for} \quad i < j, \\ \rho_{ij}\tau_2, & \text{for} \quad i > j, \end{cases} \tag{3.2}$$

where the τ parameters pertain to the upper-right triangle and lower-left triangle in the square table. Another generalization of the 0 asymmetry model is the RC asymmetry model,

$$P_{ij} = \rho_{ij}\alpha_i\beta_j, \quad \text{for} \quad i \neq j, \tag{3.3}$$

where the α_i and β_j parameters pertain to the ith row and jth column, respectively. And still another generalization is the diagonals (D) asymmetry model,

$$P_{ij} = \rho_{ij}\delta_k, \quad \text{for} \quad i \neq j, \quad k = i - j, \tag{3.4}$$

where the δ_k parameters pertain to the upper-left lower-right diagonals, with $k = i - j$, for $k = \pm 1, \pm 2, \cdots, \pm(K - 1)$, in the $K \times K$ table.

The triangle (T) asymmetry model (3.2) is related to but different from the triangles-parameter model introduced in Goodman (1972), and it is equivalent to the conditional symmetry model in Bishop *et al.* (1975, page 286), and McCullagh (1978). The RC asymmetry model (3.3) is the usual quasi-symmetry model introduced by Caussinus (1965), and it is equivalent to the symmetric association model in Goodman (1979a). And the diagonals (D) asymmetry model (3.4) is the diagonals-parameter model introduced in Goodman (1979b).

The first section in Table 16A gives the degrees of freedom pertaining to each of these asymmetry models, and the corresponding section in Table 16B gives the chi-squared values obtained when each of these models is applied to Table 3. Starting with the 0 asymmetry model (3.1) applied to Table 3, we find that this model does not fit the data adequately, and we then note that the D asymmetry model (3.4) provides a dramatic improvement in fit over the other asymmetry models considered here.

3.3. *Generalized independence models.* Each of the generalized symmetry models in Section 3.2 pertains to entries that are not on the main diagonal of the square table, and we shall consider next some other models (viz., generalized independence models) that pertain to the same entries (for $i \neq j$). First we consider the usual model of quasi-independence between the row classification and column classification:

$$P_{ij} = \alpha_i\beta_j, \quad \text{for} \quad i \neq j. \tag{3.5}$$

To facilitate the exposition in this section on generalized independence models, model (3.5) will now be called the null (0) nonindependence model (for $i \neq j$).

TABLE 16A

Degrees of freedom for generalized symmetry, independence, and symmetry + independence models applied to the $K \times K$ table (for $K \geq 3$)

models	degrees of freedom
Asymmetry models	
0	$K(K-1)/2$
T	$(K+1)(K-2)/2$
RC	$(K-1)(K-2)/2$
D	$(K-1)(K-2)/2$
Nonindependence models (with main diagonal deleted)	
0	K^2-3K+1
T	$K(K-3)$
DA	$(K-1)(K-3)$*
DAT	(K^2-4K+2)*
D	(K^2-5K+5)*
Non- symmetry + independence models (with main diagonal deleted)	
0	$K(K-2)$
T	K^2-2K-1
DA	$(K-1)(K-2)$*
DAT	(K^2-3K+1)*
D	$(K-1)(K-3)$*
Nonindependence models (with no entries deleted)	
0	$(K-1)^2$
T	K^2-2K-1
DA	$(K-1)(K-2)$
DAT	K^2-3K+1
D	$(K-2)^2$
Non- symmetry + independence models (with no entries deleted)	
0	$K(K-1)$
T	$(K+1)(K-2)$
DA	$(K-1)^2$
DAT	$K(K-2)$
D	$(K-1)(K-2)$

* The asterisk indicates that the formula does not apply for $K = 3$. For this special case, see comments in Section 3.3.

A simple generalization of model (3.5) is the triangle (T) nonindependence model,

$$P_{ij} = \begin{cases} \alpha_i\beta_j\tau_1, & \text{for } i < j, \\ \alpha_i\beta_j\tau_2, & \text{for } i > j. \end{cases} \tag{3.6}$$

In addition, we take note now of the diagonals (D), diagonals-absolute (DA), and diagonals-absolute triangle (DAT) nonindependence models:

$$P_{ij} = \alpha_i\beta_j\delta_k, \quad \text{for } i \neq j, \quad k = i - j; \tag{3.7}$$

$$P_{ij} = \alpha_i\beta_j\delta'_k, \quad \text{for } i \neq j, \quad k = |i - j|; \tag{3.8}$$

and

$$P_{ij} = \begin{cases} \alpha_i\beta_j\delta'_k\tau_1, & \text{for } i < j, \\ \alpha_i\beta_j\delta'_k\tau_2, & \text{for } i > j. \end{cases} \tag{3.9}$$

TABLE 16B
Generalized symmetry, independence, and symmetry + independence models applied to Table 3

models	degrees of freedom	goodness-of-fit chi-squared	likelihood-ratio chi-squared
Asymmetry models			
0	6	19.11	19.25
T	5	7.26	7.35
RC	3	7.26	7.27
D	3	.50	.50
Nonindependence models (with main diagonal deleted)			
0	5	198.01	199.10
T	4	195.90	196.98
DA	3	7.26	7.27
DAT	2	6.78	6.82
D	1	.22	.22
Non- symmetry + independence models (with main diagonal deleted)			
0	8	209.44	209.24
T	7	195.99	197.34
DA	6	19.11	19.25
DAT	5	7.26	7.35
D	3	.50	.50
Nonindependence models (with no entries deleted)			
0	9	8096.86	6671.50
T	7	463.03	488.12
DA	6	54.60	54.40
DAT	5	54.03	53.92
D	4	46.82	46.97

The second section in Table 16A gives the degrees of freedom pertaining to each of these generalized independence models, and the corresponding section in Table 16B gives the chi-squared values obtained when each of these models is applied to Table 3. Starting now with the 0 nonindependence model (3.5), we also find that this model does not fit the data, and we then note that the D nonindependence model (3.7) provides a dramatic improvement in fit over the other generalized independence models considered here.

Before closing this section, we note that the formulae for the degrees of freedom given in Table 16A apply to all the models considered there for $K \times K$ tables, with $K \geq 3$; except for the asterisked formulae when $K = 3$. (With respect to the models presented in this section, asterisked formulae are given in Table 16A for the DA, DAT, and D nonindependence models.) When $K = 3$, the correct formulae are obtained by adding 1 to the asterisked formulae.

3.4. *Generalized symmetry + independence models.* While the models in Section 3.2 describe various forms of asymmetry, and the models in Section 3.3 describe various forms of nonindependence (for $i \neq j$), the models that will be presented in this section can be used to study simultaneously both asymmetry and nonindependence (for $i \neq j$). First we shall consider here the model that describes simultaneously both symmetry and independence:

$$P_{ij} = \alpha_i \alpha_j, \quad \text{for} \quad i \neq j. \tag{3.10}$$

(Model (3.10) is model (3.5) with $\beta_j = \alpha_j$, for $j = 1, 2, \cdots, K$, in the $K \times K$ table.) We shall call this model the null (0) non- symmetry + independence model.

A simple generalization of model (3.10) is the triangle (T) non- symmetry + independence model

$$P_{ij} = \begin{cases} \alpha_i \alpha_j \tau_1, & \text{for } i < j, \\ \alpha_i \alpha_j \tau_2, & \text{for } i > j. \end{cases} \tag{3.11}$$

Now we take note of the diagonals (D), diagonals-absolute (DA), and diagonals-absolute triangle (DAT) non- symmetry + independence models:

$$P_{ij} = \alpha_i \alpha_j \delta_k, \quad \text{for } i \neq j, \quad k = i - j; \tag{3.12}$$

$$P_{ij} = \alpha_i \alpha_j \delta'_k, \quad \text{for } i \neq j, \quad k = |i - j|; \tag{3.13}$$

and

$$P_{ij} = \begin{cases} \alpha_i \alpha_j \delta'_k \tau_1, & \text{for } i < j, \\ \alpha_i \alpha_j \delta'_k \tau_2, & \text{for } i > j. \end{cases} \tag{3.14}$$

The null (0) non- symmetry + independence model (3.10) was described in Goodman (1981c) and Hope (1982), and the diagonals (D) non- symmetry + independence model (3.12) was introduced in Goodman (1979b, 1981c). For related material, see Sobel, Hout, and Duncan (1985).

Each of the generalized symmetry + independence models considered in the present section is included in the third section of Table 16A, and the chi-squared values obtained when these models are applied to Table 3 are included in the corresponding section of Table 16B. Starting now with the 0 non- symmetry + independence model (3.10), we also find that this model does not fit the data, and we then note that the D non- symmetry + independence model (3.12) provides a dramatic improvement in fit over the other generalized symmetry + independence models. We shall next compare the results obtained here for the generalized symmetry + independence models with those obtained in the preceding sections on generalized symmetry and generalized independence.

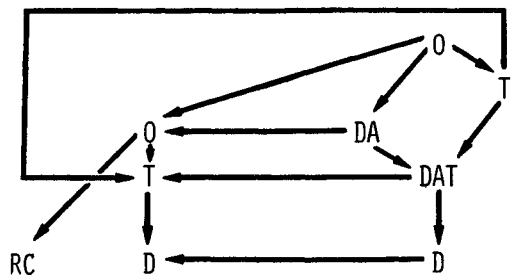
3.5. *Some comparisons, an overview, and some additional models.* From Table 16B, we saw that starting with the null model of symmetry, we were led to the D asymmetry model, which fits the data very well; starting with the null model of independence (for $i \neq j$), we were led to the D nonindependence model, which also fits the data very well; and starting with the null model of symmetry + independence, we were led to the D non- symmetry + independence model, which also fits the data very well. For the 4×4 table, using the general methods presented in Goodman (1972), we find that the D asymmetry model (3.4) and the D non- symmetry + independence model (3.12) are equivalent. Since both the D non- symmetry + independence model (3.12) and the D nonindependence model (3.7) fit the data well, we note that the former model is more parsimonious (for $K \geq 3$). (The D non- symmetry + independence model is also more parsimonious than the D asymmetry model for $K > 4$, and the two models are equivalent when $K = 3$ or 4.)

Using the general methods presented in Goodman (1972), we also find that,

for $K = 4$, the T asymmetry model (3.2) is equivalent to the DAT non- symmetry + independence model (3.14), the 0 asymmetry model (3.1) is equivalent to the DA non- symmetry + independence model (3.13), and the RC asymmetry model (3.3) is equivalent to the DA nonindependence model (3.8).

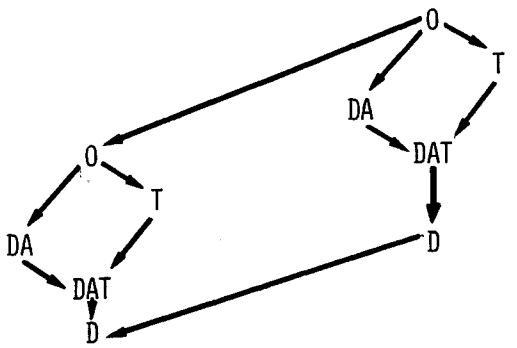
The relationships between the generalized symmetry models and generalized symmetry + independence models, with respect to their implications and degrees of freedom, are displayed in Figure 2A; and a corresponding display for the

FIG. 2A. *Generalized Symmetry and Symmetry + Independence Models*

Models		Degrees of freedom*
Symmetry	Symmetry + independence	
		8 (12) 7 (10) 6 (9) 5 (8) 3 (6)

* The degrees of freedom in parentheses are for the generalized symmetry + independence models with no entries deleted.

FIG. 2B. *Generalized Independence and Symmetry + Independence Models*

Models		Degrees of freedom*
Independence	Symmetry + independence	
		8 (12) 7 (10) 6 (9) 5 (*) 4 (7) 3 (6) 2 (5) 1 (4)

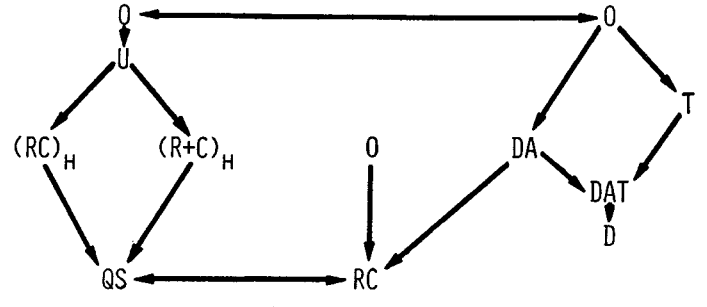
* With no entries deleted, the degrees of freedom for the 0 nonindependence model and the DAT non- symmetry + independence model are 9 and 8, respectively.

FIG. 2. *The relationship between various models with respect to their implications and degrees of freedom, for the $K \times K$ table; with degrees of freedom given for $K = 4$ with main diagonal deleted, and with no entries deleted in parentheses.*

generalized independence models and generalized symmetry + independence models is presented in Figure 2B. All the relationships described in the preceding two paragraphs are included in these figures except for the final relationship described there (viz., the relationship between the RC asymmetry model (3.3) and the DA nonindependence model), which can be inferred from the display in Figure 3.

In Sections 3.2, 3.3, and 3.4, we were concerned with models for the $K \times K$ table with the main diagonal deleted; but the models presented in the latter two sections can be directly extended to the $K \times K$ table with no entries deleted. With respect to these extensions, see the final sections in Tables 16A and 16B. In addition, we include in Figures 2 and 3 information about the generalized independence models and generalized symmetry + independence models with no entries deleted, as well as with the main diagonal deleted.

In Figure 3, we have described the relationships between the generalized independence models, some of the generalized symmetry models, and some of the association models considered in Section 2.1 above and in Goodman (1979a). All the association models in Figure 3 describe forms of symmetric association applied to the $K \times K$ table (where $I = J = K$). From (2.6) we see that the 0 association model and the U association model describe forms of symmetric association, since the association measure Φ_{ij} is equal to zero with the former model, and is constant with the latter model, for $i = 1, 2, \cdots, K - 1$, and $j = 1, 2, \cdots, K - 1$. (We have assumed in Figure 3 that the U association model is defined by (2.7)–(2.8), or by (2.9) with $\Delta_i' = \Delta_i''$, for $i = 1, 2, \cdots, K - 1$.)



* The generalized symmetry models are with main diagonal deleted, with 6 and 3 degrees of freedom for the 0 and RC models, respectively; and the QS symmetric association model is also with main diagonal deleted, and with 3 degrees of freedom. The degrees of freedom for the $(RC)_H$ and $(R + C)_H$ symmetric association models and the DA nonindependence model are 6 with no entries deleted and 3 with main diagonal deleted.

FIG. 3. *The relationship between various models with respect to their implications and degrees of freedom, for the $K \times K$ table; with degrees of freedom for $K = 4$ with no entries deleted, and with main diagonal deleted in parentheses.*

Similarly, the RC association model (2.2) with homogeneous row and column effects (i.e., the $(RC)_H$ model, with $\mu_i = \nu_i$ in (2.2), for $i = 1, 2, \cdots, K-1$), which was introduced in Goodman (1979a), and the corresponding R + C association model (2.65) with homogeneous row and column effects (i.e., the $(R + C)_H$ model, with $\alpha_i^* = \beta_i^*$ and $\Delta_i' = \Delta_i''$ in (2.65), for $i = 1, 2, \cdots, K-1$), also describe forms of symmetric association. And the DA nonindependence model (3.8) in Figure 3, which was introduced in Goodman (1972), can also be included as a form of symmetric association. Finally, we note again that the usual quasi-symmetry (QS) model (i.e., the RC asymmetry model (3.3)) is equivalent to the model of symmetric association (see, e.g., Goodman 1979a). Table 17A gives the degrees of freedom pertaining to each of these symmetric association models, and Table 17B gives the chi-squared values obtained when each of these models is applied to Table 3. Except for the usual quasi-symmetry model, the other models in Table 17A can be applied to the $K \times K$ table with no entries deleted, as well as with the main diagonal deleted. (The quasi-symmetry model can also be extended to the $K \times K$ table with no entries deleted, but we shall not consider this model here.)

TABLE 17A

Degrees of freedom for symmetric association models applied to the $K \times K$ table (for $K \geq 3$)

models	degrees of freedom
With main diagonal deleted	
0	$K^2 - 3K + 1$
U	$K(K-3)$
$(RC)_H$	$(K^2 - 4K + 2)^{**}$
$(R + C)_H$	$(K^2 - 4K + 2)^{**}$
DA	$(K-1)(K-3)^*$
QS	$(K-1)(K-2)/2$
With no entries deleted	
0	$(K-1)^2$
U	$K(K-2)$
$(RC)_H$	$(K-1)(K-2)$
$(R + C)_H$	$(K-1)(K-2)$
DA	$(K-1)(K-2)$

* The asterisks indicate that the formula does not apply for $K = 3$ (single asterisk), and for $K = 3$ and $K = 4$ (double asterisk). For this special case, see comments in Section 3.5.

TABLE 17B

Symmetric association models applied to Table 3

models	degrees of freedom	goodness-of-fit chi-squared	likelihood-ratio chi-squared
With main diagonal deleted			
0	5	198.01	199.10
U	4	64.60	61.31
$(RC)_H$	3	7.26	7.27
$(R + C)_H$	3	7.26	7.27
DA	3	7.26	7.27
QS	3	7.26	7.27

The asterisked formulae for the degrees of freedom given in Table 17A need to be interpreted with caution. With the main diagonal deleted, when $K = 3$ the $(RC)_H$ and $(R + C)_H$ models are equivalent saturated models, with 0 degrees of freedom, and the DA model has 1 degree of freedom (see Goodman, 1972); and when $K = 4$, the three models are equivalent, with 3 degrees of freedom. (Also, with the main diagonal deleted, when $K = 3$ the DA model and the QS model are equivalent, with 1 degree of freedom; and when $K = 4$, the $(RC)_H$, $(R + C)_H$, and DA models are all equivalent to the QS model, with 3 degrees of freedom.)

The class of generalized independence models considered here was introduced in Goodman (1972). Now using also the corresponding class of generalized symmetry models, generalized symmetry + independence models, and symmetric association models, which have been under consideration here, we have a more flexible model-building apparatus for the square cross-classification table. Additional generalized independence models were also included in Goodman (1972) (see, e.g., the crossings-parameter models considered there), and these models, too, can be modified in a straightforward way to provide the corresponding asymmetry, non- symmetry + independence, and symmetric association models; but we shall not go into these details here.

In discussing association models in the present section, we have considered only symmetric association models, in order to focus attention in Figure 3 on the relationship between these models and the generalized symmetry and generalized independence models. We could have also included here a more general discussion of the association and correlation models (both symmetric and asymmetric models) applied to the $K \times K$ table. Of course, these models, which were defined in Section 2 for the $I \times J$ table, can be directly applied to the $K \times K$ table (with $I = J = K$). All the association models considered in this lecture (in Sections 2 and 3) and all the generalized independence models (in Section 3) can be viewed as members of a single class of models; viz., a more general class of association models (i.e., models that do not pertain to the row marginal and column marginal of the cross-classification table). The generalized symmetry models considered in Section 3.2 (except for the RC asymmetry model (3.3)), and the generalized symmetry + independence models in Section 3.4, are not included in this larger class (i.e., the general class of association models), since they do pertain to the row marginal and column marginal of the table. (For further details, see the discussion of models of association vs. models for the joint distribution of the cross-classification table, in Goodman, 1981c.) For the sake of completeness, we also note here that, with respect to the correlation models considered in Section 2.2, the 0 and RC models are equivalent to models that are included in the general class of association models, but the U, R, and C correlation models are not; and a similar comment applies to the extensions of the correlation models in Section 2.4. (See the remarks immediately following (2.27) pertaining to the U, R, and C models.)

We comment briefly now on a model for the $K \times K$ table that is quite different from all the models considered above. This model is not included in any of the general classes of models described here, although it is somewhat related to the diagonals (D) asymmetry model (3.4). The latter model was modified earlier

herein to obtain the more parsimonious diagonals (D) non- symmetry + independence model (3.12). An alternative modification of (3.4), which also provides a more parsimonious model, is the following model:

$$P_{ij} = \delta_k \delta^*_{i+j}, \quad \text{for} \quad i \neq j, \quad k = i - j, \tag{3.15}$$

where the δ^*_{i+j} parameters pertain to the upper-right/lower-left diagonals. This model states that there is quasi-independence between the difference-diagonal classification (i.e., the difference between the row and column classification) and the sum-diagonal classification (i.e., the sum of the row and column classification), for $i \neq j$. We can refer to this quasi-independence model as $[D_d \otimes D_s]$, for $i \neq j$, where D_d and D_s denote the difference-diagonal classification and sum-diagonal classification, respectively. The number of degrees of freedom for testing this model is $K^2 - 5K + 7$ (for $K \geq 3$). The model is more parsimonious than model (3.4) for $K > 4$, and the two models are equivalent for $K = 3$ and $K = 4$. Thus, when model (3.15) is applied to Table 3, we again obtain chi-squared values of 0.50, with 3 degrees of freedom.

Instead of considering model (3.15) for $i \neq j$, we could instead have considered

$$P_{ij} = \delta_k \delta^*_{i+j}, \quad \text{for} \quad k = i - j, \tag{3.16}$$

applied to all entries in the cross-classification table. The number of degrees of freedom for this model is $(K - 2)^2$, which is greater than the corresponding number for model (3.15) except for $K = 3$, in which case the two models are equivalent. When this model is applied to Table 3, the goodness-of-fit and likelihood-ratio chi-squared values are 28.81 and 28.88, with 4 degrees of freedom. Thus, the model does not fit the data. Comparing the chi-squared values just obtained for model (3.16) with the corresponding values obtained in the preceding paragraph for model (3.15), we see that the inadequacy of fit obtained with model (3.16) is due almost entirely to the entries on the main diagonal (more particularly, to the entries in cells (2, 2) and (3, 3)).

Since (3.16) is a model of quasi-independence for the two diagonal classifications (D_d and D_s), we shall call this model the DD model. We shall also call (3.16) the diamond (DD) model, since it can be viewed as a model of quasi-independence applied to the diamond shape formed by rotating the $K \times K$ table forty-five degrees (so that the $2K - 1$ difference-diagonals in the original table now form the entries in the rows of the diamond, and the corresponding $2K - 1$ sum-diagonals in the original table form the entries in the columns of the diamond). This model is related to, but different from, a "diamond model" discussed by Hope (see, e.g., Hope, 1982). Model (3.16) is the quasi-independence model $[D_d \otimes D_s]$ with no entries deleted from the original $K \times K$ table, and model (3.15) is the quasi-independence model $[D_d \otimes D_s]$ with the main diagonal of the original $K \times K$ table deleted. In view of this, we refer to models (3.15) and (3.16) as the DD models, or as diamond models for the reason noted above.

As we noted earlier in this section, there are $(K - 2)^2$ degrees of freedom for the DD model (3.16); and as we noted in Section 2, there are also $(K - 2)^2$ degrees of freedom for the RC and R + C association (and correlation) models; see, Table 4B or Figure 1. However, the DD model (3.16) is very different from the RC and

R + C models. The DD model is a particular kind of model of quasi-independence (viz., $[D_d \otimes D_s]$), whereas the RC and R + C association (and correlation) models cannot be described in these terms.

Almost all the models for the $K \times K$ table considered in the present section (aside from the association models considered here) can be viewed simply as models of independence or quasi-independence applied to appropriate rearrangements of the cells in the table. Thus, for example, the T asymmetry model (3.2) is equivalent to the model of independence applied to a $2 \times L$ table (with $L = K(K-1)/2$) obtained by placing in the first row of the $2 \times L$ table the L entries from the upper-right triangle of the $K \times K$ table, and by placing in the second row of the $2 \times L$ table the corresponding entries from the lower-left triangle of the $K \times K$ table. Similarly, the T nonindependence model (3.6) is equivalent to the model of quasi-independence applied to a $K \times K \times 2$ table, where the first of the two layers in the three-way table is the original $K \times K$ table with the L entries from the lower-left triangle deleted, and the second layer in the three-way table is the original $K \times K$ table with the L entries from the upper-right triangle deleted. (In the three-way table with entries deleted as indicated above, we apply the model that states that the corresponding three variables—the row, column, and layer variables—are mutually quasi-independent.) As another example of this general method, the $[D_d \otimes D_s]$ model (3.16) is equivalent (as we noted earlier) to the model of quasi-independence applied to a $(2K-1) \times (2K-1)$ table obtained by rotating the $K \times K$ table 45° so that the $2K-1$ difference-diagonals form the entries in the rows of the $(2K-1) \times (2K-1)$ table and the corresponding $2K-1$ sum-diagonals form the entries in the columns of the $(2K-1) \times (2K-1)$ table. (This model can be simplified further by removing entries from the $K \times K$ table that are irrelevant with respect to this model; e.g., the entries in cells (1, 1), (K, K), $(1, K)$, and $(K, 1)$; and also by partitioning the quasi-independence model for the $(2K-1) \times (2K-1)$ table into components pertaining to models of independence applied to appropriate separable subtables.) For additional examples, see, e.g., Goodman (1972).

The only models for the $K \times K$ table considered in the present section that are not equivalent to models of independence or quasi-independence applied to appropriate rearrangements of the cells in the table are the following models: The 0 asymmetry model (3.1) (which can be viewed simply as a model of conditional equi-probability applied to each of the L columns in the $2 \times L$ table described at the beginning of the preceding paragraph); the generalized symmetry + independence models applied to the entire $K \times K$ table; and the $(RC)_H$ and $(R + C)_H$ association models. With respect to the generalized symmetry + independence models mentioned above, these models can be expressed as log-linear models, and a suitable computer program can be applied. This is true also for the $(R + C)_H$ model (and for the R + C model in Section 2); but not for the $(RC)_H$ model (nor for the RC model in Section 2). With respect to the $(RC)_H$ model (and the RC model in Section 2), and other association models as well, the methods in Goodman (1979a) are suitable.

Before closing Section 3, we comment briefly now on the application of the models considered in this section to the case where the $K \times K$ table has deleted

entries. We have already considered here the deletion of entries on the main diagonal, but we could also consider the deletion of entries in other sets of cells. As we noted above, almost all the models in this section can be viewed as models of independence or quasi-independence (in a rearrangement of the cells in the $K \times K$ table), and the few models that cannot be viewed in this way can be viewed instead as models in the more general class of log-linear models. (The only exception to the above statement is the $(RC)_H$ model in this section, which can be applied to the $K \times K$ table with or without deleted entries using the methods in Goodman, 1979a.) All the models that can be viewed as models of independence or quasi-independence can be directly extended to the case where the entries in a given set of cells are deleted (see, e.g., Goodman, 1968), and the other (log-linear) models can be extended in a similar way.

4. Unordered parameters and/or ordered parameters. Aside from the models of symmetry (3.1), quasi-symmetry (3.3), independence (3.5), symmetry + independence (3.10), and the $(RC)_H$ model, the other models in Section 3 took into account the specified order of the categories. In some contexts, it will be sufficient to consider models that take into account the specified order in the way we have done with the models of Section 3; and in other contexts, the ordering of a set of parameters in the model (as we did, say, with (2.15)) will also be worthwhile. Although we did consider briefly in Section 2 the ordering of a set of parameters in some of the models considered there (e.g., with the application of (2.15)), we have not yet considered the ordering of a set of parameters in models of the kind considered in Section 3. We shall do so now.

Let us return for a moment to, say, the D asymmetry model (3.4). The diagonals parameter δ_k in this model can be interpreted as the odds that the difference between the observed row and column classification will be equal to k rather than $-k$, for $k \geq 1$; with $\delta_k = 1$ for $k = -1, -2, \cdots, -(K-1)$. For the data in Table 3, we find that the maximum-likelihood estimates are

$$\hat{\delta}_1 = .86, \quad \hat{\delta}_2 = .99, \quad \hat{\delta}_3 = .55. \tag{4.1}$$

In view of the above interpretation of the δ_k's, we might wish to consider the following restriction on the order of the δ_k's:

$$\delta_k \geq \delta_{k+1}, \quad \text{for} \quad k = 1, 2, \cdots, K-2. \tag{4.2}$$

Since the likelihood function is concave for model (3.4) (recall that it is a log-linear model), when this model is applied to the data in Table 3 with restriction (4.2) imposed, we find that the maximum likelihood will be obtained with $\delta_1 = \delta_2$. In this case, the corresponding maximum-likelihood estimates are

$$\hat{\delta}_1 = \hat{\delta}_2 = .88, \quad \hat{\delta}_3 = .55. \tag{4.3}$$

In fitting the D asymmetry model (3.4) to the data in Table 3 (without restriction (4.2)), the goodness-of-fit and likelihood-ratio chi-squared values were both 0.50, with 3 degrees of freedom (see Table 16B); and now with the restriction $\delta_1 = \delta_2$ imposed, the corresponding chi-squared values are 2.04 and 2.03, respectively, with 4 degrees of freedom. This model also fits the data well.

In fitting the D asymmetry model (3.4), we examine the paired estimates $(\hat{\delta}_k, \hat{\delta}_{k+1})$ to see which pairs violate (4.2), for $k = 1, 2, \cdots, K - 2$. Since model (3.4) can be viewed as a log-linear model and the likelihood function is concave (Haberman, 1973), if one pair of estimates $(\hat{\delta}_k, \hat{\delta}_{k+1})$ violates (4.2) (e.g., as in (4.1), with $k = 1$), then the maximum likelihood will be obtained on the boundary determined by the one violation (e.g., with $\delta_1 = \delta_2$, when the violator is $(\hat{\delta}_1, \hat{\delta}_2)$); and so the violating pair is pooled in this case (e.g., as in (4.3), with $k = 1$). Similarly, if more than one pair violates (4.2), the maximum likelihood will be obtained with the pool-adjacent-violator rule (see, e.g., Barlow, et al., 1972, Section 2.4). To test model (3.4) with restriction (4.2) imposed, we obtain a conditional chi-squared test with $(K - 1)(K - 2)/2$ degrees of freedom (as in the case when restriction (4.2) is not imposed) if no pairs violate (4.2); and we obtain a conditional chi-squared test with $[(K - 1)(K - 2)/2] + (k - 1)$ degrees of freedom, if the pool-adjacent-violator rule yields $K - k$ distinct estimates, for $k = 1, 2, \cdots, K - 1$. (For example, for the 4×4 table (Table 3) with restriction (4.2) imposed, the chi-squared statistic obtained would have 3 degrees of freedom if no pairs violated (4.2), 4 degrees of freedom if two estimates were pooled, and 5 degrees of freedom if the three estimates were pooled.)

The conditional test considered above has the advantage of ease of application since it applies the usual tabulated chi-squared distribution, but it has the disadvantage of conditioning on a statistic (viz., the number $K - k$ of distinct estimates obtained with the pool-adjacent-violator rule) that is not ancillary (see, e.g., Cox and Hinkley, 1974, page 31). To ameliorate this disadvantage, instead of using the same level of significance α_k in assessing the conditional test with $K - k$ distinct estimates (with $\alpha_k = \alpha$, for $k = 1, 2, \cdots, K - 1$), we would suggest the use of α_k with $\alpha_k \leq \alpha_{k+1}$, for $k = 1, 2, \cdots, K - 2$.

The likelihood-ratio chi-squared statistics used in calculating the conditional test (for $k = 1, 2, \cdots, K - 1$) can also be used to form the likelihood-ratio test of the D asymmetry model (3.4) with restriction (4.2) imposed. This test is formed when the chi-squared statistic that is obtained is assessed using the percentiles of the marginal distribution of this statistic, which is (asymptotically) a mixture of $K - 1$ chi-squared distributions with degrees of freedom $[(K - 1)(K - 2)/2] + (k - 1)$, for $k = 1, 2, \cdots, K - 1$. (For related results, see, e.g., Barlow, et al., 1972, Chapter 3; Robertson and Wegman, 1978.) This mixture will depend, in general, upon the relative size of the variances of the estimates $\hat{\delta}_k$, for $k = 1, 2, \cdots, K - 1$. (When $K = 3$, there are only two estimates ($\hat{\delta}_1$ and $\hat{\delta}_2$), and the marginal distribution, in this special case, is an equal mixture of two chi-squared distributions (with 2 and 3 degrees of freedom); but for $K > 3$, the mixture of the $K - 1$ chi-squared distributions will depend upon the relative size of the variances.) If the conditional test is applied using the same level of significance $\alpha_k = \alpha$, for $k = 1, 2, \cdots, K - 1$, then this test is less conservative (i.e., more likely to reject the hypothesis when the hypothesis is true) than the corresponding test based upon the marginal distribution when the number $K - k$ of distinct estimates is equal to $K - 1$ (or is close to $K - 1$) in the conditional test; and it is more conservative than the marginal test when the number $K - k$ of distinct estimates is equal to 1 (or is close to 1) in the conditional test.

If the hypothesized restriction (4.2) is correct with strict inequality applying (for $k = 1, 2, \cdots, K - 2$), then we would find that $\hat{\delta}_k > \hat{\delta}_{k+1}$ (for $k = 1, 2, \cdots, K - 2$) almost certainly (asymptotically); and the need to pool estimates would not arise. If (4.2) is correct (for $k = 1, 2, \cdots, K - 2$) with strict inequality applying for all but one value of k (say, for $k = k'$), and $\delta_{k'} = \delta_{k'+1}$, then we would find that $\hat{\delta}_{k'} > \hat{\delta}_{k'+1}$ with probability ½, and that $\hat{\delta}_{k'} < \hat{\delta}_{k'+1}$ with probability ½ (asymptotically). In this case, we would be led to calculate the chi-squared statistic with no restrictions imposed, with probability ½; and the chi-squared statistic with one restriction imposed ($\delta_{k'} = \delta_{k'+1}$), with probability ½. (For related comments, see, e.g., Agresti, Chuang, and Kezouh, 1984.) On the other hand, if (4.2) is correct (for $k = 1, 2, \cdots, K - 2$) with strict inequality applying for all but two (or more) values of k (say, for $k = k'$ and $k = k''$, with $k' \neq k''$), then we would be led to calculate the chi-squared statistic with no restrictions imposed, or the corresponding statistic with one or more restrictions imposed (say, $\delta_{k'} = \delta_{k'+1}$ and/or $\delta_{k''} = \delta_{k''+1}$), but the corresponding probabilities associated with these events would depend upon the relative size of the variances of the $\hat{\delta}_k$.

With respect to our analysis of the data in Table 3, model (3.4) with the ordering (4.2) imposed fits the data well, if equality is applicable in the ordering; but we would reject the model with ordering (4.2) imposed if strict inequality applied in the ordering. If equality is applicable, the conditional test is based on the chi-squared value of 2.03 (or 2.04), which is assessed using the usual tabulated chi-squared distribution with 4 degrees of freedom; and the marginal test would be based on this value assessed using the appropriate mixture of three chi-squared distributions (with 3, 4, and 5 degrees of freedom). With the conditional test and also with the marginal test, we see that the model fits the data well.

We have considered above tests of the hypothesis that model (3.4) holds true with restriction (4.2) imposed. We shall now consider briefly the hypothesis that restriction (4.2) holds true under the assumption that model (3.4) holds true. To test this hypothesis, we simply consider the difference between the chi-squared statistic with restriction (4.2) imposed and the corresponding statistic without (4.2) imposed. Using this difference, we obtain a conditional chi-squared test of hypothesis (4.2), with $k - 1$ degrees of freedom, where $K - k$ is the number of distinct estimates as above (for $k = 1, 2, \cdots, K - 1$). This difference is equivalent to the corresponding test of homogeneity that is obtained when the $K - 1$ original estimates $\hat{\delta}_k$ (for $k = 1, 2, \cdots, K - 1$) are partitioned into their pooled subsets, and the original estimates within each subset are tested for homogeneity within the subset.

Consider next the hypothesis that

$$\delta_k = \delta, \quad \text{for} \quad k = 1, 2, \cdots, K - 1. \tag{4.4}$$

To test this null hypothesis under the assumption that model (3.4) holds true with restriction (4.2) imposed, we simply consider the difference between the chi-squared statistic with restriction (4.4) imposed and the corresponding statistic with (4.2) imposed. Using this difference, we obtain a conditional chi-squared test of hypothesis (4.4), with $(K - 2) - (k - 1) = K - k - 1$ degrees of freedom, where $K - k$ is the number of distinct estimates as above. This difference is

equivalent to the corresponding test of homogeneity applied to the $K - k$ distinct estimates. Tests of this general form were considered earlier by Bartholomew and others (see Barlow et al., 1972, Chapter 3), in a somewhat different, but related, context.

In the preceding two paragraphs, we considered conditional tests of hypotheses (4.2) and (4.4). The comments earlier in this section comparing conditional tests with the corresponding marginal (likelihood-ratio) tests can be applied in the present context as well. If the conditional test is used with $\alpha_k = \alpha$ (for $k = 1, 2, \cdots, K - 1$), then the conditional test of (4.2) is less conservative than the corresponding test based upon the marginal distribution, and the conditional test of (4.4) is more conservative than the corresponding test based upon the marginal distribution, when the number $K - k$ of distinct estimates is equal to $K - 1$ (or is close to $K - 1$) in the conditional test; and the reverse is the case when the number $K - k$ of distinct estimates is equal to one (or is close to one) in the conditional test.

The preceding discussion was concerned with the imposition of ordering on a set of parameters in a given model and/or the imposition of related equality restrictions on the parameters. For expository purposes, the discussion focused on the imposition of ordering (4.2) on parameters in model (3.4). The results described in this discussion are applicable more generally. For example, with respect to the imposition of the ordering (2.15) on the parameters ν_j ($j = 1, 2, \cdots, J$) in the C association model applied to Table 1 in Section 2.1, or the corresponding imposition of ordering on parameters in the analysis of Table 2 in Section 2.4, the results described above are applicable.

Since each of the models considered in Section 3 (except for the $(RC)_H$ model) is a log-linear model and the likelihood function is concave (and also, with these particular models, there is an order-preserving relationship between the maximum-likelihood estimates of the relevant parameters and the corresponding sufficient statistics), the methods described above can be applied more generally with any of these models. A similar comment applies also to the association models considered in Section 2, except for the RC model (and generalizations of it). With respect to this particular non- log-linear model, it is possible to show that its likelihood function will be concave (and the relationship between the maximum-likelihood estimates of the relevant parameters and the corresponding sufficient statistics will be order-preserving) in a neighborhood of the true value for the parameters in the model when the sample size is large, and the methods described earlier in this section can be applied with this model too (for large samples).

APPENDIX

Maximum-likelihood estimation

The method introduced in Goodman (1979a) for calculating maximum-likelihood estimates for the RC association model and related association models will be modified here for the RC correlation model and related correlation models. We shall begin with the RC correlation model (2.17).

The maximum-likelihood equations for (2.17) are

$$\sum_{i,j} p_{ij} c_{ij}(\hat{\lambda}) = 0, \tag{A.1}$$

$$\sum_j p_{ij} a_{ij}(\hat{x}_i) - p_{i\cdot} \sum_{k,j} p_{kj} a_{kj}(\hat{x}_k) = 0, \tag{A.2a}$$

$$\sum_i p_{ij} b_{ij}(\hat{y}_j) - p_{\cdot j} \sum_{i,k} p_{ik} b_{ik}(\hat{y}_k) = 0, \tag{A.2b}$$

with

$$c_{ij}(\hat{\lambda}) = \hat{x}_i \hat{y}_j / (1 + \hat{\lambda} \hat{x}_i \hat{y}_j), \tag{A.3}$$

$$a_{ij}(\hat{x}_i) = \hat{\lambda} \hat{y}_j / (1 + \hat{\lambda} \hat{x}_i \hat{y}_j), \tag{A.4a}$$

$$b_{ij}(\hat{y}_j) = \hat{\lambda} \hat{x}_i / (1 + \hat{\lambda} \hat{x}_i \hat{y}_j). \tag{A.4b}$$

For the RC model, formula (2.27) also holds true. For simplicity, we assume that $p_{i\cdot} > 0$ and $p_{\cdot j} > 0$. In addition, when $\hat{\lambda} = 0$, the $\hat{x}_i$ and $\hat{y}_j$ are unidentifiable, and so we also assume here that $\hat{\lambda} \neq 0$ (i.e., that $p_{ij} \neq p_{i\cdot} p_{\cdot j}$ for some i and j).

To solve (A.1)–(A.2), we can apply Newton's elementary (unidimensional) method, thus obtaining the following formulae which can be applied iteratively:

$$\lambda^{**} = \lambda^* + \sum_{i,j} p_{ij} c_{ij}(\lambda^*) / \sum_{i,j} p_{ij} [c_{ij}(\lambda^*)]^2, \tag{A.5}$$

$$x_i^{**} = x_i^* + \frac{\sum_j p_{ij} a_{ij}(x_i^*) - p_{i\cdot} \sum_{k,j} p_{kj} a_{kj}(x_k^*)}{\sum_j p_{ij} [a_{ij}(x_i^*)]^2 (1 - p_{i\cdot})}, \tag{A.6a}$$

$$y_j^{**} = y_j^* + \frac{\sum_i p_{ij} b_{ij}(y_j^*) - p_{\cdot j} \sum_{i,k} p_{ik} b_{ik}(y_k^*)}{\sum_i p_{ij} [b_{ij}(y_j^*)]^2 (1 - p_{\cdot j})}. \tag{A.6b}$$

From (2.27), we also obtain

$$\sum_i \hat{x}_i p_{i\cdot} = 0, \quad \sum_j \hat{y}_j p_{\cdot j} = 0, \tag{A.7}$$

$$\sum_i \hat{x}_i^2 p_{i\cdot} = 1, \quad \sum_j \hat{y}_j^2 p_{\cdot j} = 1; \tag{A.8}$$

and the trial values for $\hat{x}_i$ and $\hat{y}_j$ can be adjusted at each iterative step to satisfy (A.7)–(A.8). (It actually is not necessary to make the adjustment to satisfy (A.8) at each iterative step, since the purpose of this adjustment is only to define the scale of the $\hat{x}_i$ and $\hat{y}_j$; and this can be done at the end of the iterations.)

Note that (A.2a) implies (A.1) if $\sum_i \hat{x}_i p_{i\cdot} = 0$, and (A.2b) implies (A.1) if $\sum_j \hat{y}_j p_{\cdot j} = 0$. Thus, equation (A.1) is redundant; and so we can delete (A.5), and the $\hat{\lambda}$ term in (A.4), and the adjustment to satisfy (A.8) for either the $\hat{x}_i$ or the $\hat{y}_j$. If the adjustment to satisfy (A.8) is deleted for the $\hat{x}_i$, we now define $\hat{x}_i$ to be the product of the original $\hat{\lambda}$ and the original $\hat{x}_i$. With initial trial values for $\hat{x}_i$ (as now defined) and $\hat{y}_j$ inserted in (A.6a), this formula can be applied to obtain a new trial value for $\hat{x}_i$, then this new value can be used in (A.6b) to obtain a new trial value for $\hat{y}_j$, then the new values of $\hat{x}_i$ and $\hat{y}_j$ can be used again in (A.6a). Et cetera.

The iterative method described above for the RC model (2.17)—i.e., (2.56) with $M^* = 1$—can be directly extended to the more general case of (2.56) with $1 \leq M^* \leq M$. Initial trial values can be obtained from the analysis of the saturated

model with $M = \min(I - 1, J - 1)$, using the singular value decomposition (see, e.g., Kendall and Stuart, 1979, page 599); but in come cases more elementary methods can be used to obtain initial values. (For example, estimates obtained below with the U correlation model can be used sometimes to suggest initial trial values for the R and C correlation models, and estimates obtained below with the R and C correlation models can be used sometimes to suggest initial trial values for the RC model.)

The iterative method described above, using Newton's elementary (unidimensional) method, is different from the scoring algorithm applied in Gilula and Haberman (1984). However, the iterative method and the scoring algorithm will have similar convergence properties. The iterative method is easier to apply, but the rate of convergence will be slower.

We next consider the U correlation model (2.17) with (2.24) imposed. For simplicity, we begin with the estimation of the corresponding "approximate correlation model" in which restriction (2.27) is also imposed. In this case, (A.1) is applicable, (A.2) can be deleted, and the $\hat{x}_i$ and $\hat{y}_j$ can be expressed explicitly. With

$$\text{(A.9)} \qquad \hat{x}_i' = -(\textstyle\sum_{k=1}^{i-1} \Delta_k^*) + \hat{x}', \quad \hat{y}_j' = -(\sum_{k=1}^{j-1} \Delta_k^{**}) + \hat{y}',$$

$$\text{(A.10)} \qquad \hat{x}' = \textstyle\sum_i p_{i.}(\sum_{k=1}^{i-1} \Delta_k^*), \qquad \hat{y}' = \sum_j p_{.j}(\sum_{k=1}^{j-1} \Delta_k^{**}),$$

the corresponding $\hat{x}_i$ and $\hat{y}_j$ are obtained by imposing (A.8). With $\hat{x}_i$ and $\hat{y}_j$ thus defined, (A.1) can be solved iteratively using (A.5). (The initial trial value for λ^* in (A.5) can be, say, the correlation between $\hat{x}_i$ and $\hat{y}_j$ in the observed $I \times J$ table.) This simple procedure provides the estimates for the "approximate correlation model"; but it needs to be modified to obtain the maximum-likelihood estimates for the U correlation model (2.17) with (2.24) imposed (without the imposition of (2.27)). (The approximate model will serve as a good approximation when $|\lambda|$ is small (e.g., as in Table 2), but it will be less good when $|\lambda|$ is larger (e.g., as in Table 1).) We next show how to modify the simple procedure described above.

With the U correlation model, we will now find it simpler not to impose (A.8), and the $\hat{x}_i'$ and $\hat{y}_j'$ defined by (A.9) will be used directly, with (A.10) replaced by

$$\text{(A.11)} \qquad x' = \textstyle\sum_i \hat{P}_{i.}(\sum_{k=1}^{i-1} \Delta_k^*), \quad y' = \sum_j \hat{P}_{.j}(\sum_{k=1}^{j-1} \Delta_k^{**}).$$

With $\hat{x}_i'$ and $\hat{y}_j'$ thus defined, we shall now define the parameter λ' to be the ratio of the original λ and $\{[\sum_i (\hat{x}_i')^2 \hat{P}_{i.}][\sum_j (\hat{y}_j')^2 \hat{P}_{.j}]\}^{1/2}$. The maximum likelihood equations for the U correlation model are

$$\text{(A.12)} \qquad \textstyle\sum_{i,j} p_{ij} \hat{x}_i' \hat{y}_j' / (1 + \hat{\lambda}' \hat{x}_i' \hat{y}_j') = 0,$$

with

$$\text{(A.13)} \qquad \hat{P}_{i.} = p_{i.}/(1 + \hat{x}_i' a), \quad \hat{P}_{.j} = p_{.j}/(1 + \hat{y}_j' b),$$

where

$$\text{(A.14)} \quad a = \hat{\lambda}' \textstyle\sum_{i,j} p_{ij} \hat{y}_j' / (1 + \hat{\lambda}' \hat{x}_i' \hat{y}_j'), \quad b = \hat{\lambda}' \sum_{i,j} p_{ij} \hat{x}_i' / (1 + \hat{\lambda}' \hat{x}_i' \hat{y}_j').$$

From (A.9), (A.11), and (A.13), we see that

$$\sum_i \hat{P}_{i.} = 1, \quad \sum_j \hat{P}_{.j} = 1; \tag{A.15}$$

and from (A.13) and (A.15), we find that

$$a = (\sum_i \hat{x}_i' p_{i.})/\sum_i (\hat{x}_i')^2 \hat{P}_{i.}, \quad b = (\sum_j \hat{y}_j' p_{.j})/\sum_j (\hat{y}_j')^2 \hat{P}_{.j}. \tag{A.16}$$

Note that (A.12) is of the same form as (A.1), and it can be solved in the same way as (A.1), using (A.5) with the $\hat{x}_i$, $\hat{y}_j$, and $\hat{\lambda}$ replaced by $\hat{x}_i'$, $\hat{y}_j'$, and $\hat{\lambda}'$, respectively. With initial trial values for $\hat{x}_i'$, $\hat{y}_j'$, and $\hat{\lambda}'$ inserted in (A.5), this formula can be applied to obtain a new trial value for $\hat{\lambda}'$, then this new value can be used in (A.14) to calculate trial values for a and b, which can be used in turn in (A.13) to obtain trial values for $\hat{P}_{i.}$ and $\hat{P}_{.j}$, which can be used in turn in (A.9) and (A.11) to obtain new values for $\hat{x}_i'$ and $\hat{y}_j'$; then the new values of $\hat{x}_i'$, $\hat{y}_j'$, and $\hat{\lambda}'$ can be inserted again in (A.5). Et cetera. (With $\hat{x}_i'$, $\hat{y}_j'$, and $\hat{\lambda}'$ as defined above, the intial trial value of $\hat{\lambda}'$ can be, say, the ratio of the correlation between $\hat{x}_i'$ and $\hat{y}_j'$ in the observed $I \times J$ table and $\{[\sum_i (\hat{x}_i')^2 \hat{P}_{i.}][\sum_j (\hat{y}_j')^2 \hat{P}_{.j}]\}^{1/2}$, with $p_{i.}$ and $p_{.j}$ taken as initial trial values for $\hat{P}_{i.}$ and $\hat{P}_{.j}$.)

Next we comment briefly on the R and C correlation models (2.17) with (2.25) and (2.26) imposed, respectively. With respect to the estimation of the corresponding "approximate correlation models" in which restriction (2.27) is also imposed, a suitable set of formulae can be obtained by combining the appropriate part of (A.9)–(A.10) for the U approximate correlation model with the appropriate part of the maximum-likelihood equations for the RC model (see (A.1)–(A.3)). With respect to the maximum-likelihood estimation of the R and C correlation models (without the imposition of (2.27)), the corresponding equations can be obtained by combining the appropriate part of (A.11)–(A.15) for the U correlation model with the appropriate part of the maximum-likelihood equations for the RC model. However, formulae (A.2a) and (A.2b) should be replaced by

$$\sum_j p_{ij} a_{ij}(\hat{x}_i) - \hat{P}_{i.} \sum_{k,j} p_{kj} a_{kj}(\hat{x}_k) = 0, \tag{A.17a}$$

$$\sum_i p_{ij} b_{ij}(\hat{y}_j) - \hat{P}_{.j} \sum_{i,k} p_{ik} b_{ik}(\hat{y}_k) = 0; \tag{A.17b}$$

and (A.6) would also need to be modified appropriately.

We shall now use the above formulae to prove that (2.27) holds true for the RC model, that $\hat{P}_{.j} = p_{.j}$ for the R model, and that $\hat{P}_{i.} = p_{i.}$ for the C model. Under the RC model, both (A.17) and (A.13)–(A.14) hold true (with the prime superscripts deleted from the latter formulae). Thus, (A.17) can be rewritten as

$$\hat{\lambda}\hat{x}_i \sum_j p_{ij}\hat{y}_j/(1 + \hat{\lambda}\hat{x}_i\hat{y}_j) - \hat{P}_{i.}\hat{x}_i a = 0, \tag{A.18a}$$

$$\hat{\lambda}\hat{y}_j \sum_i p_{ij}\hat{x}_i/(1 + \hat{\lambda}\hat{x}_i\hat{y}_j) - \hat{P}_{.j}\hat{y}_j b = 0, \tag{A.18b}$$

which is equivalent to

$$p_{i.} - \sum_j p_{ij}/(1 + \hat{\lambda}\hat{x}_i\hat{y}_j) - \hat{P}_{i.}\hat{x}_i a = 0, \tag{A.19a}$$

$$p_{.j} - \sum_i p_{ij}/(1 + \hat{\lambda}\hat{x}_i\hat{y}_j) - \hat{P}_{.j}\hat{y}_j b = 0. \tag{A.19b}$$

From (A.13) and (A.19), we see that

$$\hat{P}_{i.} = \sum_j p_{ij} /(1 + \hat{\lambda}\hat{x}_i\hat{y}_j), \tag{A.20a}$$

$$\hat{P}_{.j} = \sum_i p_{ij}/(1 + \hat{\lambda}\hat{x}_i\hat{y}_j). \tag{A.20b}$$

From (A.20), we find that the condition that $\sum_i \hat{x}_i\hat{P}_{i.} = 0$ and $\sum_j \hat{y}_j\hat{P}_{.j} = 0$ is equivalent to the condition that

$$\sum_{ij} p_{ij}\hat{x}_i/(1 + \hat{\lambda}\hat{x}_i\hat{y}_j) = 0, \tag{A.21a}$$

$$\sum_{i,j} p_{ij}\hat{y}_j/(1 + \hat{\lambda}\hat{x}_i\hat{y}_j) = 0; \tag{A.21b}$$

so that b and a defined by (A.14) are equal to zero. Thus, from (A.13) we see that $\hat{P}_{.j} = p_{.j}$ and $\hat{P}_{i.} = p_{i.}$ for the RC model. □

For the R correlation model, the $\hat{y}_j'$ are defined by (A.9) and the $\hat{y}_j$ are defined accordingly, while the $\hat{x}_i$ are obtained from (A.17a). Thus, we obtain (A.18a), (A.19a), (A.20a), and (A.21a), which proves that b defined by (A.14) is equal to zero; and therefore, from the second equation in (A.13), we also see that $\hat{P}_{.j} = p_{.j}$. Similarly, for the C correlation model, we find that $\hat{P}_{i.} = p_{i.}$.

It is also worth noting that, when $|\hat{\lambda}'|$ is close to zero, then both a and b in (A.14) are close to zero (when the summation term on the right-hand side of each equation in (A.14) is not too large). Thus, when $|\hat{\lambda}'|$ is close to zero, we find that $\hat{P}_{i.}$ is approximately equal to $p_{i.}$ with the R correlation model, and that $\hat{P}_{.j}$ is approximately equal to $p_{.j}$ with the C correlation model. (In the preceding paragraph, we had noted that $\hat{P}_{.j} = p_{.j}$ with the R model, and that $\hat{P}_{i.} = p_{i.}$ with the C model.)

Before closing this section, we also take note of the fact that, since the likelihood function for these models need not be concave throughout the parameter space, the iterative method introduced here should use initial trial values that are in a neighborhood of the true value for the parameters. Care should be exercised to insure that a global maximum (rather than a local maximum) has been obtained for the likelihood function. Various sets of initial trial values can be used, and also various alternative versions of the iterative method can be applied. (For example, with the iterative method described above for the RC model, a new trial value was obtained first for $\hat{x}_i$, then for $\hat{y}_j$, then again for $\hat{x}_i$; et cetera. And with an alternative version of the iterative method, a new trial value could be obtained first for $\hat{y}_j$, then for $\hat{x}_i$, then again for $\hat{y}_j$; et cetera.) The comments in this paragraph apply also to the other particular non- log-linear models discussed in this article (e.g., the RC association model and generalizations of it).

REFERENCES

AGRESTI, A. (1984). *Analysis of Ordinal Categorical Data.* Wiley, New York.

AGRESTI, A., CHUANG, C., and KEZOUH, A. (1984). Order-restricted score parameters in association models for contingency tables. Manuscript.

AITKEN, A. C., and GONIN, H. T. (1935). Fourfold sampling with and without replacement. *Proc. Roy. Soc. Edinburgh* **55** 114–125.

AITKIN, M. (1982). Review of *Analysis of Categorical Data: Dual Scaling and Its Applications*, by

S. Nishisato. *J. Roy. Statist. Soc. A* **145** 513–516.

ANDERSEN, E. B. (1980). *Discrete Statistical Models with Social Science Applications.* North-Holland, Amsterdam.

ANSCOMBE, F. J. (1981). *Computing in Statistical Science through APL.* Springer-Verlag, New York.

BARLOW, R. E., BARTHOLOMEW, D. J., BREMNER, J. M., and BRUNK, H. D. (1972). *Statistical Inference Under Order Restrictions.* Wiley, New York.

BENZÉCRI, J.-P., ET AL. (1973). *L'Analyse des Données, Vol. 1: La Taxinomie, Vol. 2: L'Analyse des Correspondances.* Dunod, Paris (Second Edition 1976).

BIRCH, M. W. (1965). The detection of partial association. II: The general case. *J. Roy. Statist. Soc. Ser. B* **27** 111–124.

BISHOP, Y., FIENBERG, S. E., and HOLLAND, P. W. (1975). *Discrete Multivariate Analysis.* M.I.T. Press, Cambridge.

CAUSSINUS, H. (1965). Contribution à l'analyse statistique des tableaux de corrélation. *Annales de la Faculté des Sciences de l'Université de Toulouse* **29** 77–182.

CHUANG, C., GHEVA, D., and ODOROFF, C. (1983). Methods for diagnosing multiplicative-interaction models for two-way contingency tables. Manuscript.

CLOGG, C. C. (1982a). Using association models in sociological research: some examples. *Amer. J. Sociology* **88** 114–134.

CLOGG, C. C. (1982b). Some models for the analysis of association in multiway cross-classifications having ordered categories. *J. Amer. Statist. Assoc.* **77** 803–815.

COX, D. R., and HINKLEY, D. V. (1974). *Theoretical Statistics.* Chapman and Hall, London.

DEVILLE, J.-C., and MALINVAUD, E. (1983). Data analysis in official socio-economic statistics. *J. Roy. Statist. Soc. A* **146** 335–361.

DUNCAN, O. D. (1979). How destination depends on origin in the occupational mobility table. *Amer. J. Sociology* **84** 793–803.

ESCOUFIER, Y. (1982). L'analyse des tableaux de contingence simples et multiples. *Proc. Internat. Meeting on the Analysis of Multidimensional Contingency Tables* (Rome, 1981). R. Coppi (editor). *Metron* **40** 53–77.

GABRIEL, K. R. (1971). The biplot graphic display of matrices with applications to principal components analysis. *Biometrika* **58** 453–467.

GILULA, Z. (1984). Grouping and association in two-way contingency tables: a canonical correlation analytic approach. Manuscript.

GILULA, Z., and HABERMAN, S. (1984). Canonical analysis of contingency tables by maximum likelihood. Manuscript.

GOKHALE, D. V. and KULLBACK, S. (1978). *The Information in Contingency Tables.* Dekker, New York.

GOODMAN, L. A. (1968). The analysis of cross-classified data: independence, quasi-independence, and interactions in contingency tables with or without missing entries. *J. Amer. Statist. Assoc.* **63** 1091–1131.

GOODMAN, L. A. (1972). Some multiplicative models for the analysis of cross-classified data. *Proc. Sixth Berk. Symp.*, L. Le Cam et al. (editors). University of California Press, Berkeley.

GOODMAN, L. A. (1973). Guided and unguided methods for the selection of models for a set of T multidimensional contingency tables. *J. Amer. Statist. Assoc.* **68** 165–175.

GOODMAN, L. A. (1978). *Analyzing Qualitative/Categorical Data: Log-Linear Models and Latent-Structure Analysis.* Abt Books, Cambridge, Massachusetts.

GOODMAN, L. A. (1979a). Simple models for the analysis of association in cross-classifications having ordered categories. *J. Amer. Statist. Assoc.* **74** 537–552.

GOODMAN, L. A. (1979b). Multiplicative models for square contingency tables with ordered categories. *Biometrika* **66** 413–418.

GOODMAN, L. A. (1979c). Multiplicative models for the analysis of occupational mobility tables and other kinds of cross-classification tables. *Amer. J. Sociology* **84** 804–819.

GOODMAN, L. A. (1981a). Association models and canonical correlation in the analysis of cross-classifications having ordered categories. *J. Amer. Statist. Assoc.* **76** 320–334.

GOODMAN, L. A. (1981b). Association models and the bivariate normal for contingency tables with ordered categories. *Biometrika* **68** 347–355.

GOODMAN, L. A. (1981c). Three elementary views of log-linear models for the analysis of cross-

classifications having ordered categories. *Sociological Methodology 1981*. S. Leinhardt (editor). Jossey-Bass, San Francisco.

GOODMAN, L. A. (1981d). Criteria for determining whether certain categories in a cross-classification table should be combined, with special reference to occupational categories in an occupational mobility table. *Amer. J. Sociology* **87** 612–650.

GOODMAN, L. A. (1982). Discussion in the Round Table on "The comparison of different approaches to the analysis of contingency tables, in the light of applications." *Proc. Internat. Meeting on the Analysis of Multidimensional Contingency Tables* (Rome, 1981). R. Coppi (editor). *Metron* **40** 344–352.

GOODMAN, L. A. (1984). *The Analysis of Cross-Classified Data Having Ordered Categories*. Harvard University Press, Cambridge.

GOODMAN, L. A., and HABERMAN, S. J. (1984). On the analysis of non-additivity in two-way analysis of variance. Manuscript.

GOODMAN, L. A., and KRUSKAL, W. H. (1954). Measures of association for cross classifications. *J. Amer. Statist. Assoc.* **49** 732–764.

GOODMAN, L. A., and KRUSKAL, W. H. (1979). *Measures of Association for Cross Classifications*. Springer-Verlag, New York.

GREENACRE, M. J. (1984). *Correspondence Analysis*. Academic Press, New York.

GROSS, S. T. (1981). On asymptotic power and efficiency of tests of independence in contingency tables with ordered classifications. *J. Amer. Statist. Assoc.* **76** 935–941.

HABERMAN, S. J. (1973). Log-linear models for frequency data: sufficient statistics and likelihood equations. *Ann. Statist.* **1** 617–632.

HABERMAN, S. J. (1974a). Log-linear models for frequency tables with ordered classifications. *Biometrics* **30** 589–600.

HABERMAN, S. J. (1974b). *The Analysis of Frequency Data*. University of Chicago Press, Chicago.

HABERMAN, S. J. (1979). *Analysis of Qualitative Data: Vol. 2 New Developments*. Academic Press, New York.

HABERMAN, S. J. (1981). Tests for independence in two-way contingency tables based on canonical correlation and on linear-by-linear interaction. *Ann. Statist.* **9** 1178–1186.

HAMDAN, M. A., and MARTINSON, E. O. (1971). Maximum likelihood estimation in the bivariate binomial (0, 1) distribution: application to 2×2 tables. *Austral. J. Statist.* **13** 154–158.

HILL, M. O. (1974). Correspondence analysis: a neglected multivariate method. *Appl. Statist.* **23** 340–354.

HOPE, K. (1982). Vertical and non-vertical mobility in three countries. *Amer. Sociol. Rev.* **47** 99–113.

JOHNSON, D. E., and GRAYBILL, F. A. (1972). An analysis of a two-way model with interaction and no replication. *J. Amer. Statist. Assoc.* **67** 862–868.

KENDALL, M. G., and STUART, A. (1979). *The Advanced Theory of Statistics*, Vol. 2, Fourth Edition. Hafner, New York.

KRUSKAL, W. H. (1958). Ordinal measures of association. *J. Amer. Statist. Assoc.* **53** 814–861.

LANCASTER, H. O. (1957). Some properties of the bivariate normal distribution considered in the form of a contingency table. *Biometrika* **44** 289–292.

LANCASTER, H. O. (1969). *The Chi-Squared Distribution*. Wiley, New York.

LEBART, L., MORINEAU, A., and WARWICK, K. M. (1984). *Multivariate Descriptive Statistical Analysis*. Wiley, New York.

MANDEL, J. (1971). A new analysis of variance model for non-additive data. *Technometrics* **13** 1–18.

MARDIA, K. V. (1967). Some contributions to contingency-type bivariate distributions. *Biometrika* **54** 235–249; corr. **55** 597.

MARDIA, K. V. (1970). *Families of Bivariate Distributions*. Hafner, Darien, Connecticut.

MCCULLAGH, P. (1978). A class of parametric models for the analysis of square contingency tables with ordered categories. *Biometrika* **65** 413–418.

MOSTELLER, F. (1968). Association and estimation in contingency tables. *J. Amer. Statist. Assoc.* **63** 1–28.

NISHISATO, S. (1980). *Analysis of Categorical Data: Dual Scaling and Its Applications*. University of Toronto Press, Toronto.

O'NEILL, M. E. (1978). Distributional expansions for canonical correlations from contingency tables. *J. Roy. Statist. Soc. B* **40** 303–312.

PEARSON, K. (1913). Notes on the surface of constant association. *Biometrika* **9** 534–537.

PLACKETT, R. L. (1965). A class of bivariate distributions. *J. Amer. Statist. Assoc.* **60** 516–522.

PLACKETT, R. L. (1981). *The Analysis of Categorical Data*, Second Edition. Macmillan, New York.

RASCH, G. (1966). *Statistikkens Teori.* Lecture notes compiled by U. Christiansen in Danish. University of Copenhagen.

RIETZ, H. L. (1912). On the theory of correlation with special reference to certain significant loci on the plane of distribution in the case of normal correlation. *Ann. Math.* **13** 187–199.

RIETZ, H. L. (1916). On the correlation of marks in mathematics and law. *J. Educ. Psych.* **7** 87–92.

RIETZ, H. L. (1919). On functional relations for which the coefficient of correlation is zero. *Quart. Pub. Amer. Stat. Assn.* **16** 472–476.

RIETZ, H. L. (1920). Urn schemata as a basis for the development of correlation theory. *Ann. Math.* **21** 306–322.

RIETZ, H. L. (editor-in-chief) (1924). *Handbook of Mathematical Statistics.* Houghton Mifflin, Boston.

RIETZ, H. L. (1927). *Mathematical Statistics*, Carus Mathematical Monographs, Number Three. Mathematical Association of America. Open Court Publishing Co., Chicago. (Reprinted by George Banta Publishing Co., Menosha, Wisconsin.)

RIETZ, H. L. (1932). A simple non-normal correlation surface. *Biometrika* **24** 288–290.

RIETZ, H. L., and SHADE, I. (1908). Correlation of efficiency in mathematics and efficiency in other subjects—a statistical study. *Illinois University Studies* **3** 19–39.

RIETZ, H. L., and SMITH, L. H. (1910). *On the Measurement of Correlation with Special Reference to Some Characteristics of Indian Corn.* Bulletin No. 148, University of Illinois Agricultural Experimental Station, Urbana, Illinois.

ROBERTSON, T., and WEGMAN, E. J. (1978). Likelihood ratio tests for order restrictions in exponential families. *Ann. Statist.* **6** 485–505.

SIMON, G. (1974). Alternative analyses for the singly-ordered contingency table. *J. Amer. Statist. Assoc.* **69** 971–976.

SOBEL, M. E., HOUT, M., and DUNCAN, O. D. (1985). Exchange, structure, and symmetry in occupational mobility. To appear in *Amer. J. Sociology* **91**.

SROLE, L., LANGNER, T. S., MICHAEL, S. T., OPLER, M. K., and RENNIE, T. A. C. (1962). *Mental Health in the Metropolis: The Midtown Manhattan Study.* McGraw-Hill, New York.

STUART, A. (1953). The estimation and comparison of strengths of association in contingency tables. *Biometrika* **40** 105–110.

STUART, A. (1955). A test for homogeneity of the marginal distributions in a two-way classification. *Biometrika* **42** 412–416.

TRELOAR, A. E. (1939). *Elements of Statistical Reasoning.* Wiley, New York.

TUKEY, J. W. (1977). *Exploratory Data Analysis.* Addison-Wesley, Reading, Massachusetts.

WILKS, S. S. (1931). On the distribution of statistics in samples from a normal population of two variables with matched sampling of one variable. Ph.D. dissertation. University of Iowa, Iowa City.

WILKS, S. S. (1962). *Mathematical Statistics.* Wiley, New York.

WILLIAMS, E. J. (1952). Use of scores for the analysis of association in contingency tables. *Biometrika* **39** 274–280.

YATES, F. (1948). The analysis of contingency tables with groupings based on quantitative characters. *Biometrika* **35** 176–181; corr. **35** 424.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CHICAGO
CHICAGO, ILLINOIS 60637