A General Linear Approach to the Analysis of Nonmetric Data: Applications for Political Science

Author(s): Robert G. Lehnen and Gary G. Koch
Source: *American Journal of Political Science*, Vol. 18, No. 2 (May, 1974), pp. 283-313
Published by: Midwest Political Science Association
Stable URL: https://www.jstor.org/stable/2110703
Accessed: 08-01-2020 16:12 UTC

ROBERT G. LEHNEN
University of Houston

GARY G. KOCH
University of North Carolina

*A General Linear Approach to the Analysis of Nonmetric Data: Applications for Political Science**

A general linear approach to the analysis of nonmetric (nominal and/or ordinal) data developed for problems common to the health sciences is extended to the field of political science. After a brief description of the method originally presented by Grizzle, Starmer, and Koch (*Biometrics,* 1969), two examples are discussed in detail. The first example, using an ordered dependent variable, illustrates an analysis of variance without assumptions of normality. The data are from the University of Michigan 1964 Presidential Election Study. The second example, based on data about the disposition of petty criminal court cases in North Carolina, involves an application where the independent and dependent variables are nominal.

The period since 1945 has brought substantial changes in the methodology of political scientists. As the profession has borrowed or developed techniques suited to its problems, detailed and complex analyses of several variable problems have replaced relatively simple descriptive methods. Thus, analysis of variance, regression and correlation analysis, causal models, principal components, and factor analysis are examples of techniques that have become familiar and useful tools for many political scientists. In spite of the advancing expertise in the discipline, however, a number of persistent problems have repeatedly appeared in analysis that have compromised conclusions or invalidated results.

The most persistent of these problems is that throughout the various

*An earlier version of this paper was presented at the 1972 annual meeting of the Midwest Political Science Association in Chicago, Illinois.

Part-time support for Robert G. Lehnen was provided by a grant from the Carolina Population Center and the Department of Political Science, University of North Carolina; support for Gary G. Koch was provided by a Public Health Service Research Career Development Award (No. GM-70004) from the National Institute of General Medical Sciences. We wish to thank Dr. David Brady, Department of Political Science, University of Houston, for his helpful criticisms of prior drafts, and Mrs. Julia Hutcherson and Mrs. Pat Sanford for their assistance in typing the manuscript.

subfields of political science, the level of measurement attained is "low"—nominal or ordinal. Many of the standard statistical methods, however, require continuous data or, more specifically, data measured on an interval scale.[1] Although several scaling approaches permit the "improvement" of the level of measurement, the corresponding assumptions are not always appropriate. Moreover, some important types of data are inherently qualitative. Such measurement problems permeate the fields of political science, sociology, psychology, and many of the health sciences.

A second difficulty which is frequently encountered in the analysis of political data is the inflexibility of many statistical approaches in the face of certain design problems. Data may be generated from a broad variety of sampling schemes that sometimes lead to such familiar difficulties as disproportionate sampling and missing or incomplete data. In many Survey Research Center election studies, for example, the sample is stratified by race; yet the full advantages presented by such a design are often negated or ignored by the use of reweighting schemes. A further difficulty is that data for certain subpopulations (*e.g.*, black college-educated Republicans) are exceedingly sparse or absent altogether. Such design problems are not always readily handled by the standard methodologies.

A final problem causing difficulty for the researcher attempting to formulate empirically based models is the lack of sufficient statistical distribution theory to permit hypothesis testing as a method of inference related to the use of a given analytical technique. Even for situations where the first two problems can be handled and estimation of parameters is possible in a descriptive context, one is many times limited to either intuitive or heuristic criteria for making decisions about data.

In this paper, we shall present an alternative analytical strategy which provides a solution to the previously suggested problems of measurement, design, and testing. The method is based on the approach of Grizzle, Starmer, and Koch (hereafter abbreviated as GSK) to the analysis of complex contingency tables.[2] After a brief description of the theory underlying this approach, we shall illustrate its utility in terms of a number of political science examples.

Before turning to the description of the methodology and the applications, several alternative approaches to nonmetric data deserve some comment. Depending on the configuration of the variables, the level of measurement of

[1] The term "categorical data" is also used in reference to data that are either nominal or ordinal.

[2] J.E. Grizzle, C.F. Starmer, and G.G. Koch, "Analysis of Categorical Data by Linear Models," *Biometrics,* 25 (September 1969), 489–504.

the dependent variables, and the hypotheses of interest, both dummy variable regression and analysis-of-variance approaches have been applied to categorical data situations. In general, the application of these two techniques gives little more than limited heuristic and descriptive information. Neither provides a means for statistical inferences. Analysis-of-variance approaches require unwarranted assumptions both about the underlying distributions and about homogeneity of variance. Where the dependent variable is nominal or ordinal, such assumptions become even more tenuous. In our first example, we shall illustrate, among other things, an analysis-of-variance type of application that does not require these assumptions.

Dummy variable regression, beyond some descriptive value, also does not provide a satisfactory alternative. As with analysis of variance, no suitable testing procedures are available, and applications involving an ordinally scaled dependent variable are suspect. Both techniques fail, however, where the dependent variable is nominally scaled. The second example illustrates an application of GSK where all variables including the dependent variable are nominal.

Finally, neither technique provides a clear criterion for accepting or rejecting the linear model fitted to the data. Both analysis of variance and dummy variable regression involve an implicit linear model, but the ambiguity associated with decisions about goodness of fit have caused such considerations generally to be ignored. The GSK approach provides a statistically grounded goodness-of-fit (GOF) test of the linear model that is a potentially valuable tool for improving and verifying theory-building efforts.

Analysis of Complex Contingency Tables

The Standard Chi-Square Approach

Nonmetric or "categorical" data are usually presented in the form of multidimensional contingency tables that have been derived from multiway cross-tabulations. The cell entries in these tables are the frequencies of occurrence for phenomena corresponding to the respective combinations of the variables under study. These contingency tables may be regarded in a general manner with certain variables having purely nominal values or levels and others having ordinally scaled values. If intervally scaled variables are grouped into appropriate class intervals, they also can be treated within this same framework.

Until recently, the information that could be obtained from such contingency table data has been limited to estimates for the proportions of the various nonmetric response categories together with certain χ^2 statistics which have been calculated for various two-dimensional subtables. In many situations, such χ^2 statistics are directed at hypotheses which are fragmented into interrelated parts. Certain complex hypotheses that are of potentially greater interest, such as those pertaining to the main effects and interactions in multidimensional contingency tables, are not directly testable by the usual χ^2 methods.[3]

Thus, if one were interested in the effects of race, income, and education on party identification, traditional χ^2 methods do not provide a way to test the significance of each effect on party preferences, nor do they provide any unified method for studying the interaction of race, income, and education on party preferences. The most frequent solution has been to analyze three two-dimensional tables—one each for race, income, and education—by party identification, while ignoring hypotheses about interactions. Since the hypotheses about the main effects are interrelated, any approach that seeks to estimate parameters and test hypotheses needs to consider the complexities of these interrelationships.

The GSK Approach

One way of dealing with many of these analytical difficulties is to use the approach described by Grizzle, Starmer, and Koch. This methodology is based on the application of general weighted least squares to estimates of appropriate functions of the cell proportions in the complex contingency table. From these results, valid statistical tests of complex hypotheses of interest are produced.

The basic assumption required by the GSK approach is that the set of estimated proportions in a complex contingency table has approximately a multivariate normal distribution with a covariance matrix that can be consistently estimated in a satisfactory manner. Hence, whenever the data under study have arisen from a probability random sample survey design with known probabilities of selection, this methodology is directly applicable, provided that the sample size is sufficiently large.[4] In these cases, estimates for the covariance matrix can be based on standard principles of sampling

[3] V.P. Bhapkar and G.G. Koch, "On the Hypothesis of 'No Interaction' in Contingency Tables," *Biometrics,* 24 (February 1968), 567–94; and Bhapkar and Koch, "Hypothesis of No Interaction in Multidimensional Contingency Tables," *Technometrics,* 10 (February 1968), 107–22.

[4] Each cell proportion should be based on at least 10 observations and ideally 25 or more; for certain rare attributes, as many as 100 observations may be required.

theory as discussed in Cochran, or on replication methods as discussed in McCarthy or Kish and Frankel.[5] Alternatively, "retrospective" data–that is, data that have arisen from an *ad hoc* sampling design–may also be analyzed in terms of this approach provided that one can argue that the data could have conceivably arisen from an underlying product multinomial distribution model. In other words, such retrospective data must be assumed to be equivalent possibly to a stratified simple random sample from an infinite population. Most data used in political science studies meet one of these two assumptions: either the data come from a more or less rigorous sampling design (sample surveys), or "nature" provides the sampling and the researcher can accept assumptions about their representativeness with respect to some defined population to permit some mode of statistical analysis (*e.g.* many elite studies).

With the previous remarks in mind, let us now consider the hypothetical data in Table 1. The rows of this array correspond to separate subpopulations (or strata) from which independent random samples of *a priori* fixed sizes n_1, $n_2, \ldots, n_s$ have been obtained. The subjects in each of these samples have been classified into categories according to some generalized set of responses, so that n_{ij} represents the estimated (weighted) frequency of occurrence of the j-th response category in the sample from the i-th subpopulation.[6] In this

TABLE 1

Hypothetical Data and Corresponding Probabilities of Occurrence

	Response Category				
Subpopulation	j = 1	2	...	r	Total
i = 1	n_{11} (p_{11})	n_{12} (p_{12})	...	n_{1r} (p_{1r})	n_1
2	n_{21} (p_{21})	n_{22} (p_{22})	...	n_{2r} (p_{2r})	n_2
...	...	...	...	...	...
s	n_{s1} (p_{s1})	n_{s2} (p_{s2})	...	n_{sr} (p_{sr})	n_s

$p_{ij} = (n_{ij}/n_i)$ $\qquad$ $\sum_{j=1}^{r} p_{ij} = 1.00$

[5] W. G. Cochran, *Sampling Techniques,* (New York: John Wiley & Sons, 1963, 2nd ed.); P.J. McCarthy, "Pseudo-Replication: Half Samples," *Review of the International Statistical Institute,* 37 (1969), 239–64; and L. Kish and M.R. Frankel, "Balanced Repeated Replication for Analytical Statistics," in American Statistical Association, *Proceedings, Social Statistics Section,* (1968), pp. 2–10.

[6] The phrase "weighted frequency" pertains to situations where the data have been obtained from complex sample surveys and reflect adjustments accounting for unequal probabilities of selection.

framework, both i and j can be complex—*e.g.*, the subpopulations could represent combinations of two or more factors, like region of the country and place of residence, and the responses could represent combinations of opinions pertaining to agreement or disagreement with two or more policy questions. For example, the values of i and j could conceivably be as in Table 2 where A, B, and C represent three policy questions. In this context, n_{44} represents the number of individuals in the sample from the Northcentral-Rural region who agree with statement A and disagree with both B and C.

From the data in Table 1, the quantities $p_{ij} = (n_{ij}/n_i)$ can be determined as estimates for the probability of occurrence of the j-th response category in the i-th subpopulation. In order to maintain simplicity in exposition, we shall assume that the elements of the covariance matrix of the $\{p_{ij}\}$ can be consistently estimated by expressions of the type

$$\mathrm{Var}\{p_{ij}\} = \frac{p_{ij}\,(1-p_{ij})}{n_i}, \quad \mathrm{Cov}\{p_{ij}, p_{ij'}\} = \frac{-p_{ij}p_{ij'}}{n_i} \quad \text{where } j \neq j' \tag{1.1}$$

$$\mathrm{Cov}\{p_{ij}, p_{i'j'}\} = 0 \quad \text{where } i \neq i' \tag{1.2}$$

with Eq. 1.2 being a consequence of the independence of the samples from the different subpopulations. This point of view is entirely valid if the data have been obtained from independent simple random samples[7] or from retrospective samples where the product multinomial probability model is

TABLE 2

Example of Complex Subpopulation and Response Categories

Subpopulation Categories (Region-Residence)		Response Categories
Northeast	Urban : i = 1	Agree A, Agree B, Agree C : j = 1
Northeast	Rural : i = 2	Agree A, Agree B, Disagree C : j = 2
Northcentral	Urban : i = 3	Agree A, Disagree B, Agree C : j = 3
Northcentral	Rural : i = 4	Agree A, Disagree B, Disagree C : j = 4
South	Urban : i = 5	Disagree A, Agree B, Agree C : j = 5
South	Rural : i = 6	Disagree A, Agree B, Disagree C : j = 6
West	Urban : i = 7	Disagree A, Disagree B, Agree C : j = 7
West	Rural : i = 8	Disagree A, Disagree B, Disagree C : j = 8

[7] If the simple random samples had been selected without replacement from small or moderate size finite populations, one may adjust the results in Eq. 1.3 by the necessary finite population correction. See W.D. Johnson and G.G. Koch, "Analysis of Qualitative Data: Linear Functions," *Health Services Research,* 5 (Winter 1970), 358–69.

applicable. Otherwise, it is important to recognize that when the data have arisen from complex sample survey designs, then Eq. 1.1 no longer applies even in an approximate sense. On the other hand, the emphasis in this paper will be placed on those aspects of an analysis which pertain to relationships among variables and/or have a multiple regression flavor. For this type of investigation, certain empirical results of Kish and Frankel suggest that the complex sample survey design effect is small and that Eq. 1.1 can be used as a reasonable approximation in the preliminary stages of analysis.[8] In the remainder of this section, matrix notation and arithmetic will be used to summarize some of the mathematical aspects of the GSK approach relevant to the examples to be discussed later. Readers with interest in the more theoretical details should refer to the Appendix.

Let the vectors $\mathbf{n}_i$ and $\mathbf{p}_i$ and the matrix $\mathbf{V}_i$ correspond to the arrays (the rows of Table 1) as defined in Eq. 1.3 where

$$\underset{(r\times 1)}{\mathbf{n}_i} = \begin{bmatrix} n_{i1} \\ n_{i2} \\ \cdots \\ n_{ir} \end{bmatrix}, \quad \underset{(r\times 1)}{\mathbf{p}_i} = \begin{bmatrix} p_{i1} \\ p_{i2} \\ \cdots \\ p_{ir} \end{bmatrix}, \quad \underset{(r\times r)}{\mathbf{V}_i} = \frac{1}{n_i}\begin{bmatrix} p_{i1}(1-p_{i1}) & -p_{i1}p_{i2} & \cdots & -p_{i1}p_{ir} \\ -p_{i1}p_{i2} & p_{i2}(1-p_{i2}) & \cdots & -p_{i2}p_{ir} \\ \cdots & \cdots & \cdots & \cdots \\ -p_{i1}p_{ir} & -p_{i2}p_{ir} & \cdots & p_{ir}(1-p_{ir}) \end{bmatrix} \tag{1.3}$$

$i = 1, 2, ..., s$. Similarly, the composite vectors $\mathbf{n}$ and $\mathbf{p}$ and the composite matrix $\mathbf{V}$ correspond to the arrays defined in Eq. 1.4 where $\mathbf{0}$ is an $(r \times r)$ matrix

$$\underset{(rs\times 1)}{\mathbf{n}} = \begin{bmatrix} \mathbf{n}_1 \\ \mathbf{n}_2 \\ \cdots \\ \mathbf{n}_s \end{bmatrix}, \quad \underset{(rs\times 1)}{\mathbf{p}} = \begin{bmatrix} \mathbf{p}_1 \\ \mathbf{p}_2 \\ \cdots \\ \mathbf{p}_s \end{bmatrix}, \quad \underset{(rs\times rs)}{\mathbf{V}} = \begin{bmatrix} \mathbf{V}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_2 & \cdots & \mathbf{0} \\ \cdots & \cdots & \cdots & \cdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{V}_s \end{bmatrix} \tag{1.4}$$

of 0's. In this notation, $\mathbf{V}$ represents the estimated covariance matrix for the vector $\mathbf{p}$ of estimated probabilities. Taking the entries p_{ij} from Table 1 as an example, $\mathbf{p}'$ is a vector of probabilities derived from the frequency data n_{ij} arrayed in Table 1.

$$\underset{(1\times rs)}{\mathbf{p}'} = [\, p_{11}\, , p_{12}\, , \cdots p_{1r},\ p_{21},\ p_{22}, \cdots\ p_{2r},\ \ldots p_{sr}\,] \tag{1.5}$$

$\mathbf{V}$ is the variance-covariance matrix associated with $\mathbf{p}$.

[8] Kish and Frankel, "Balanced Repeated Replication for Analytical Statistics"; and M.R. Frankel, *Inference From Survey Samples: An Empirical Investigation,* (Ann Arbor: Institute for Social Research, University of Michigan, 1971).

Most research questions involve the study of selected functions based on **p**. Two broad classes of functions analyzable by the GSK approach are the linear and log-linear functions of **p**. Using matrix notation, a set of u linear functions can be written in the form

$$\underset{(u \times 1)}{\mathbf{a}} = (a_\xi) = \left(\sum_{i=1}^{s} \sum_{j=1}^{r} a_{\xi,ij} p_{ij} \right) = \underset{(u \times rs)}{\mathbf{A}} \; \underset{(rs \times 1)}{\mathbf{p}} \tag{1.6}$$

where $\mathbf{A} = (a_{\xi,ij})$ is a known $(u \times rs)$ matrix. Similarly, a set of m log-linear functions can be written in the form

$$\begin{aligned} \underset{(m \times 1)}{\mathbf{f}} &= (f_\eta) = \left(\sum_{\xi=1}^{u} k_{\eta\xi} \left\{ \log_e [a_\xi] \right\} \right) = \left(\sum_{\xi=1}^{u} k_{\eta\xi} \left\{ \log_e \left[\sum_{i=1}^{s} \sum_{j=1}^{r} a_{\xi,ij} p_{ij} \right] \right\} \right) \\ &= \underset{(m \times u)}{\mathbf{K}} \left[\mathbf{log_e} \left\{ \underset{(u \times rs)}{\mathbf{A}} \; \underset{(rs \times 1)}{\mathbf{p}} \right\} \right] = \underset{(m \times u)}{\mathbf{K}} \left[\mathbf{log_e} \; \underset{(u \times 1)}{\mathbf{a}} \right] \end{aligned} \tag{1.7}$$

where $\mathbf{K} = (k_{\eta,\xi})$ is a known $(m \times u)$ matrix and $\mathbf{log_e}$ transforms the **a** vector to the corresponding vector of natural logarithms.

Using the hypothetical probabilities presented in Table 3,

$$\underset{(1 \times 8)}{\mathbf{p}'} = [\ .40, .60, .20, .80, .35, .65, .75, .25\ ].$$

A possible set of u linear functions (see Eq. 1.6), the probability of agreement for each subpopulation, is defined as follows:

$$\underset{(4 \times 1)}{\mathbf{a}} = \underset{(4 \times 8)}{\mathbf{A}} \; \underset{(8 \times 1)}{\mathbf{p}} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \times \underset{(8 \times 1)}{\mathbf{p}} = \begin{bmatrix} .40 \\ .20 \\ .35 \\ .75 \end{bmatrix} \tag{1.8}$$

A set of m log-linear functions—say, the log ratio of agreement to disagreement—may be defined by the following **A** and **K** matrices:

$$\underset{(8\times 8)}{\mathbf{A}} = \mathbf{I}_8 = \begin{bmatrix} 1&0&0&0&0&0&0&0\\ 0&1&0&0&0&0&0&0\\ 0&0&1&0&0&0&0&0\\ 0&0&0&1&0&0&0&0\\ 0&0&0&0&1&0&0&0\\ 0&0&0&0&0&1&0&0\\ 0&0&0&0&0&0&1&0\\ 0&0&0&0&0&0&0&1 \end{bmatrix} \quad \underset{(4\times 8)}{\mathbf{K}} = \begin{bmatrix} 1&-1&0&0&0&0&0&0\\ 0&0&1&-1&0&0&0&0\\ 0&0&0&0&1&-1&0&0\\ 0&0&0&0&0&0&1&-1 \end{bmatrix} \qquad (1.9)$$

Using the expression in Eq. 1.7, one has

$$\underset{(4\times 1)}{\mathbf{f}} = \underset{(4\times 8)}{\mathbf{K}} \left[\log_e \left\{ \underset{(8\times 8)}{\mathbf{A}} \underset{(8\times 1)}{\mathbf{p}} \right\} \right] = \begin{bmatrix} \log_e(.40) - \log_e(.60) \\ \log_e(.20) - \log_e(.80) \\ \log_e(.35) - \log_e(.65) \\ \log_e(.75) - \log_e(.25) \end{bmatrix} \qquad (1.10)$$

By the proper choice of the **A** and **K** matrices, one may generate many different kinds of functions depending upon the structure of the particular data in a given situation. A discussion of the formulation and use of other functions is given in Forthofer and Koch.[9]

TABLE 3

Hypothetical Probabilities of Response

		Response Categories j=1	2	
Subpopulation		Agree	Disagree	Total
A and B	i = 1	.40	.60	1.00
A and not B	2	.20	.80	1.00
not A and B	3	.35	.65	1.00
not A and not B	4	.75	.25	1.00

r = 2; s = 4; rs = 8

[9] R.N. Forthofer and G.G. Koch, "An Analysis for Compounded Functions of Categorical Data," *Biometrics*, 29 (March 1973), 143–57.

The variation in the elements of either the vector **a** or the vector **f** can be analyzed by fitting linear regression models by the method of weighted least squares. Weighted least squares, unlike conventional least squares procedures, does not require assumptions of equal variance (homoscedasticity). In this context let **F** denote an appropriate set of g functions–that is, **F** is a vector either of linear functions **a** or of log-linear functions **f**. A linear model of the type

$$\underset{(g\times 1)}{\mathbf{F}} = (F_\gamma) \mathrel{\hat{=}} \left(\sum_{\delta=1}^{t} x_{\gamma\delta} b_\delta\right) = \underset{(g\times t)}{\mathbf{X}}\ \underset{(t\times 1)}{\mathbf{b}} \tag{1.11}$$

may be fitted to the data where $\mathbf{X} = (x_{\gamma\delta})$ is a $(g\times t)$ known coefficient matrix of full rank $t \leqslant g$, **b** is a $(t \times 1)$ vector of estimated parameters, and "$\hat{=}$" means "is estimated by." The definition of the design matrix **X** depends on the nature of the sample data, the functions defined on **p**, and the hypotheses of interest. The application of **X** will be illustrated in some detail in the examples.

At this point, it is important to note that a goodness-of-fit statistic is provided by the GSK approach for assessing how well the model **Xb** characterizes the functions **F**. If the model does adequately characterize the data, the difference between **F** and the values predicted by the model will be small, and the GOF Chi-square statistic will be nonsignificant. If the GOF Chi-square statistic is significant, the discrepancies between observed and predicted values (residuals) are sufficiently large to warrant fitting an alternative model **X**.

Given that the model has an acceptable fit, tests of hypotheses about the parameters estimated from the data are possible. The general hypothesis is of the form

$$H_0:\ \underset{(d\times t)}{\mathbf{C}}\ \underset{(t\times 1)}{\mathbf{b}} = \underset{(d\times 1)}{\mathbf{0}} \tag{1.12}$$

where **b** is the vector of parameters estimated for the model and **C** is a matrix selecting the parameters to be tested equal to zero. Several illustrations of the use of **C** will be given in the examples.

Examples

Mean Scores in Opinion Surveys: An Application Involving Weighted Analysis of Variance of Linear Functions

For data collected in opinion surveys, the researcher often wants to study the effects of certain characteristics of the respondent on a particular attitudinal response. Here, we shall consider an example based on the 1964 Presidential Election Survey conducted by the University of Michigan Survey Research Center. The data, which are shown in Table 4, represent frequencies of the responses "strongly agree," "agree," "ambivalent," "disagree," and "strongly disagree" to the question

TABLE 4

Attitudes Toward School Desegregation by Race X Education X Political Party

			Attitude Toward School Integration						
			Strongly Agree	Agree	Ambiv-alent	Dis-agree	Strongly Disagree	Estimated Average Score	Esti-mated Variance
Race	Edu-cation	Party	(2)	(1)	(0)	(−1)	(−2)		
B	LHS	D	180	5	7	2	14	1.61	0.0056
B	LHS	I	10	0	1	0	2	1.23	0.1675
B	LHS	R	9	1	0	0	0	1.90	0.0090
B	HS	D	44	3	2	0	2	1.71	0.0148
B	HS	I	2	0	1	0	0	1.33	0.2963
B	HS	R	1	0	0	0	0	2.00	0.0000
B	SC	D	46	0	3	1	4	1.54	0.0252
B	SC	I	3	1	1	0	1	0.83	0.3565
B	SC	R	5	0	2	0	0	1.43	0.1166
W	LHS	D	108	11	33	15	154	−0.30	0.0101
W	LHS	I	13	0	2	2	20	−0.43	0.0928
W	LHS	R	36	3	10	4	65	−0.50	0.0274
W	HS	D	69	9	12	8	63	0.08	0.0210
W	HS	I	9	1	0	1	8	0.11	0.1933
W	HS	R	32	5	7	2	45	−0.25	0.0373
W	SC	D	103	13	17	4	81	0.24	0.0156
W	SC	I	14	1	3	1	18	−0.22	0.0937
W	SC	R	83	8	20	7	75	0.09	0.0173

Some people say that the government in Washington should see to it that white and Negro (colored) children are allowed to go to the same schools. Others claim that this is not the government's business. Have you been concerned enough about this question to favor one side over the other?

for individuals from subpopulations corresponding to combinations of the following three factors:[10]

Factors	*Factor Categories*
Race	Black (B), White (W)
Education	Less Than High School (LHS), High School Diploma (HS), Some College (SC)
Political Party	Democrat (D), Independent (I), Republican (R)

Since the principal problem of interest is to investigate the degree of intensity of response for the opinion holders in each of the eighteen Race × Education × Party (2 × 3 × 3) subpopulations, respondents classified as having "no opinion" are excluded from this discussion.

One way to summarize the opinions of the subjects in each of the subpopulations is to assign scores to the response categories and then to obtain estimated average scores. In assigning these scores (or weights) a researcher probably would wish to characterize the valence and intensity of response to the school desegregation question. One possible set of scores is 2, 1, 0, −1, −2, corresponding to "strongly agree," "agree," "ambivalent," "disagree," and "strongly disagree." Then the appropriate "average opinion scores" for each subpopulation can be formulated in terms of Eq. 1.6 with

$$\underset{18 \times 90}{\mathbf{A}} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{0}' & \dots & \mathbf{0}' \\ \mathbf{0}' & \mathbf{A}_1 & \dots & \mathbf{0}' \\ \dots & \dots & \dots & \dots \\ \mathbf{0}' & \mathbf{0}' & \dots & \mathbf{A}_1 \end{bmatrix} \quad \text{where } \mathbf{A}_1 = [2 \;\; 1 \;\; 0 \;\; {-1} \;\; {-2}] \qquad (2.1)$$

[10] The choice as to which variables are "factors" and which are "responses" usually depends on the objectives of the analysis. Here, race must be treated as a factor because the data include the oversample of blacks provided by the sample design. Education and party have been chosen as additional factors because of their theoretical interest as possible determinants of opinion.

and where $\mathbf{0}'$ is a (1×5) row vector of 0's. These estimates together with corresponding estimated variances based on Eq. A.1 are given in the last two columns of Table 4.

The second phase of the analysis is concerned with the formulation of a model which reflects the extent to which the "average opinion scores" are influenced by race, education, and party. As we have seen, this can be done by using weighted least squares in the manner described in the Appendix, provided that the sample sizes in each of the subpopulations are sufficiently large. Here, however, there are several factor categories in which the number of respondents is really too small, even though the oversample of blacks has been used without reweighting. With respect to analyses involving average scores, "too small" means less than 5 cases per subpopulation. These subpopulations are black Independents and black Republicans who have either a high school diploma or some college.

TABLE 5
Observed and Predicted "Average Opinion Scores" on School Desegregation Question for Modified Race × Education × Party Groups

Race	Education	Party	Estimated Average Score	Estimated Variance of Average Score	Predicted (X_2) Average Score	Predicted Variance
B	LHS	D	1.61	0.0056	1.62	0.0035
B	LHS	I	1.23	0.1675	1.12	0.0906
B	LHS	R	1.90	0.0090	1.87	0.0082
B	HS	D	1.71	0.0148	1.62	0.0035
B	HS or SC	I	1.00	0.1975	1.12	0.0906
B	HS or SC	R	1.50	0.0938	1.87	0.0082
B	SC	D	1.54	0.0252	1.62	0.0035
W	LHS	D	−0.30	0.0101	−0.27	0.0073
W	LHS	I	−0.43	0.0928	−0.49	0.0116
W	LHS	R	−0.50	0.0274	−0.49	0.0116
W	HS	D	0.08	0.0210	0.00	0.0049
W	HS	I	0.11	0.1933	−0.22	0.0069
W	HS	R	−0.25	0.0373	−0.22	0.0069
W	SC	D	0.24	0.0156	0.27	0.0098
W	SC	I	−0.22	0.0937	0.05	0.0095
W	SC	R	0.09	0.0173	0.05	0.0095

One way of bypassing this difficulty is to pool the data for certain subpopulations that can be assumed to be reasonably similar and then to fit models to the correspondingly modified set of "average opinion scores." This strategy will be employed in this analysis by combining the High School Diploma and Some College categories each for black Independents and black Republicans. The revised set of subpopulations, together with corresponding "average opinion scores" and estimated variances are given in Table 5. These were obtained by applying Equations 1.6 and A.1 for an **A** matrix like that in Eq. 2.1 but with dimensions (16×80) instead of (18×90). In order to obtain an initial sense of the relative importance of the main effects and interactions of the Race, Education, and Party factors, one often fits a complete factorial model like that specified by the matrix $\mathbf{X}_1$ in Eq. 2.2. The rows of $\mathbf{X}_1$ may be identified with the respective subpopulations in Table 5, and the columns of $\mathbf{X}_1$ may be identified with the respective sources

$$\mathbf{X}_1 = \left[\begin{array}{rrrrrrrrrrrrrrrr}
1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 1 & 1 \\
1 & 1 & 1 & 0 & 0 & -1 & 1 & 0 & 0 & -1 & 0 & -1 & 0 & 0 & 0 & -1 \\
1 & 1 & 1 & 0 & -1 & 0 & 1 & 0 & -1 & 0 & -1 & 0 & 0 & 0 & -1 & 0 \\
1 & 1 & -1 & 1 & 1 & 1 & -1 & 1 & 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\
1 & 1 & -1 & -.33 & 0 & -1 & -1 & -.33 & 0 & -1 & 0 & 1 & 0 & .33 & 0 & 1 \\
1 & 1 & -1 & -.75 & -1 & 0 & -1 & -.75 & -1 & 0 & 1 & 0 & .75 & 0 & 1 & 0 \\
1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 \\
\hdashline
1 & -1 & 1 & 0 & 1 & 1 & -1 & 0 & -1 & -1 & 1 & 1 & 0 & 0 & -1 & -1 \\
1 & -1 & 1 & 0 & 0 & -1 & -1 & 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 & 1 \\
1 & -1 & 1 & 0 & -1 & 0 & -1 & 0 & 1 & 0 & -1 & 0 & 0 & 0 & 1 & 0 \\
1 & -1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 \\
1 & -1 & -1 & 1 & 0 & -1 & 1 & -1 & 0 & 1 & 0 & 1 & 0 & -1 & 0 & -1 \\
1 & -1 & -1 & 1 & -1 & 0 & 1 & -1 & 1 & 0 & 1 & 0 & -1 & 0 & -1 & 0 \\
1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 & 1 & 1 \\
1 & -1 & -1 & -1 & 0 & -1 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & -1 \\
1 & -1 & -1 & -1 & -1 & 0 & 1 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & -1 & 0
\end{array}\right] \tag{2.2}$$

of variation given in Tables 6 or 7. The method of construction used to derive $\mathbf{X}_1$ is primarily based on indicator (or dummy variable) functions of the type shown in Table 6. However, the rows of $\mathbf{X}_1$ associated with the pooled subpopulations require special considerations reflecting the relative numbers of respondents in the subpopulations that comprise them. For black Independents, (3/9) are HS and (6/9) are SC; while for black Republicans, (1/8) are HS and (7/8) are SC. These proportions have no influence on the

TABLE 6

Analysis of Variance for Saturated Attitude Model

Source of Variation	Columns of **X**	Estimated Parameters	Indicator Variables	D. F.	X^2
Overall Mean	1	b_1	$x_1 = 1$ always	–	–
Main Effect of Race	2	b_2	$x_2 = -1$ if B; -1 if W	1	143.86
Main Effect of Education	3,4	b_3, b_4	$x_3 =$ 1 if LHS; -1 if HS; -1 if SC, $x_4 =$ 0 if LHS; 1 if HS; -1 if SC	2	0.78
Main Effect of Party	5,6	b_5, b_6	$x_5 =$ 1 if D; 0 if I; -1 if R, $x_6 =$ 1 if D; -1 if I; 0 if R	2	2.78
Race × Education Interaction	7,8	b_7, b_8	$x_7 = x_2 x_3$, $x_8 = x_2 x_4$	2	6.81
Race × Party Interaction	9,10	b_9, b_{10}	$x_9 = x_2 x_5$, $x_{10} = x_2 x_6$	2	2.76
Education × Party Interaction	11,12,13,14	$b_{11}, b_{12}, b_{13}, b_{14}$	$x_{11} = x_3 x_5$, $x_{12} = x_3 x_6$, $x_{13} = x_4 x_5$, $x_{14} = x_4 x_6$	4	2.22
Race × Education × Party Interaction	15,16	b_{15}, b_{16}	$x_{15} = x_2 x_3 x_5$, $x_{16} = x_2 x_3 x_6$	2	0.72
Total	2–15	$b_2, b_3, \ldots b_{16}$	——	15	646.17

variables which have the same value for both HS and SC, since in these cases they are added with the sum being unity. However, with the variables x_4, x_8, x_{13}, and x_{14}, HS and SC have values of opposite sign (See Table 6). Hence, the appropriate coefficients for black Independents are $(3/9)(\pm 1) + (6/9)(\mp 1) = (\mp 0.33)$, while those for black Republicans are $(1/8)(\pm 1) + (7/8)(\mp 1) = (\mp 0.75)$. When these adjustments are made in the definition of $\mathbf{X}_1$, the only loss of information from the pooling strategy employed occurs for the Race X Education X Party interaction, where only two variables instead of four are available for estimation purposes.

Statistical tests of significance for the estimated parameters corresponding to the sources of variation listed in Table 6 can be undertaken by applying Eq. A.6 with the appropriate choices of **C** matrices. For example, to test the Education X Party interaction, one can use **C** as shown in Eq. 2.3.

$$\mathbf{C} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix} \qquad (2.3)$$

These results, which are also given in Table 6, indicate that race is the most significant factor with the Race X Education interaction also being noteworthy. Although the other effects appear to be unimportant, some caution should be exercised before drawing final conclusions. The occurrence of moderately large sample sizes for certain factor combinations and very small sample sizes for others introduces a possibly high degree of intercorrelation among the estimates of the corresponding parameters. An indication that such a problem, sometimes called "multicollinearity," exists in these data is suggested by the X^2 for the "Total" source of variation with D.F. = 15, which is almost four times as large as the sum of the separate X^2's for each component.

For this reason, an "incomplete hierarchical" model with race treated as the dominant factor was investigated. This type of model, which will be discussed in more detail in succeeding pages, is useful for characterizing interactions by reflecting the extent to which factors like party and education have different effects for blacks as opposed to whites. Moreover, it provides a basis for constructing a final model showing the most important sources of variation. With these data, such a model is specified by $\mathbf{X}_2$, defined in Eq. 2.4. Statistical tests of significance for the effects in $\mathbf{X}_2$ as well as corresponding estimated parameters are given in Table 7.

TABLE 7

Analysis of Variance for Final Attitude Model

Source of Variation	Parameters	D.F.	X^2	Estimate	Standard Error
Mean	b_1	—	—	0.71	0.060
Race	b_2	1	190.50	0.82	0.060
Party for Blacks	$\left\{\begin{matrix} b_3 \\ b_4 \end{matrix}\right.$	2	8.50	$\left\{\begin{matrix} -0.33 \\ 0.41 \end{matrix}\right.$	0.119 0.204
Education for Whites	b_5	1	19.70	−0.27	0.061
Party for Whites (D vs. [R or I])	b_6	1	3.85	0.11	0.054
Overall Model		5	641.79	—	—
Residual		10	4.38	—	—
Total		15	646.17	—	—

$$
\mathbf{X}_2 = \begin{bmatrix}
1 & 1 & 1 & 1 & 0 & 0 \\
1 & 1 & 0 & -1 & 0 & 0 \\
1 & 1 & -1 & 0 & 0 & 0 \\
1 & 1 & 1 & 1 & 0 & 0 \\
1 & 1 & 0 & -1 & 0 & 0 \\
1 & 1 & -1 & 0 & 0 & 0 \\
1 & 1 & 1 & 1 & 0 & 0 \\
\hline
1 & -1 & 0 & 0 & 1 & 1 \\
1 & -1 & 0 & 0 & 1 & -1 \\
1 & -1 & 0 & 0 & 1 & -1 \\
1 & -1 & 0 & 0 & 0 & 1 \\
1 & -1 & 0 & 0 & 0 & -1 \\
1 & -1 & 0 & 0 & 0 & -1 \\
1 & -1 & 0 & 0 & -1 & 1 \\
1 & -1 & 0 & 0 & -1 & -1 \\
1 & -1 & 0 & 0 & -1 & -1
\end{bmatrix} \tag{2.4}
$$

It can be noted that the Residual X^2 statistic in Table 7 is definitely nonsignificant (α=.25). Hence, one can conclude that the model based on $\mathbf{X}_2$ provides a satisfactory explanation of the way in which the "average opinion scores" are influenced by race, education, and party effects. Moreover, each of the estimated parameters in this model is statistically significant ($\alpha = .05$), with race, as previously noted, being the most important factor.

Predicted values for the "average opinion scores," which have been determined by applying Eq. A.7 to the estimated parameters in Table 7, are given in Table 5. The estimated variances for these quantities, which are based on Eq. A.8, are also given there, and are uniformly smaller than those corresponding to the estimates which were initially formulated in terms of Eq. 1.6.

The final attitude model for the school desegregation question shows an anticipated difference between blacks and whites, the latter being less favorable toward the issue than the former (See Table 7, parameter b_2). Education had a positive linear effect for whites—that is, attitudes became more favorable as education increased (b_5); education did not affect the attitudes of blacks. Finally, political party had different effects for each race. Among blacks, Republicans were the most favorable, followed next by Democrats and then by Independents (b_3 and b_4). Among whites, Democrats had more supportive attitudes than both Republicans and Independents (b_6).

Thus, the methodology described earlier has been used to produce a nearly complete analysis of the data in Table 4. In particular, it has been possible to study the "average opinion scores" and to use valid tests of significance to identify the nature of the effects of the race, education, and party factors and their interactions. These results provided the motivation for the formulation of a simplified model that characterized the important sources of variation and that could be used to obtain improved estimates. As a result, it has been possible to analyze this aspect of categorical data in a manner similar to traditional analysis of variance methods for continuous data. The primary difference here, however, was first that the possibly different variances of the respective "average opinion scores" were taken into proper account (unlike analysis of variance methods) by the $\mathbf{V}_a$ matrix based on Eq. A.1 and by its role in Eq. A.3–A.8. Secondly, "interactions" were analyzed in a statistically appropriate and substantively meaningful way by applying the notion of the "hierarchical" model, Eq. 2.4. Finally, some design problems associated with inadequate data were handled by making modest but acceptable assumptions, pooling the data, and designing a model (Eq. 2.2) to meet the inadequacies of the data.

Defendant Behavior in Criminal Court Proceedings: An Application Involving Multivariate Log-Linear Functions

With the emergence of the "law and order" issue in the late 1960s, the subfield of public law and judicial behavior has focused more attention on the examination of the criminal and civil court processes, especially at the local level. In this regard, we shall consider an example drawn from a larger study of the District Courts of the General Court of Justice for the State of North Carolina, which was undertaken by the Governor's Committee on Law and Order.[11] Some implications of this investigation of the distribution, detection, and disposition of criminal offenses for the criminal court process have been discussed by Lehnen and Williams.[12] The data in Table 8, which pertain to two counties, were obtained from separate independent simple random samples with replacement from the corresponding criminal court dockets.[13] The response or outcome of primary interest is the manner in which the defendant's case is disposed—*i.e.,* the case is either: brought to trial and the defendant pleads guilty (G); brought to trial and the defendant pleads not guilty (NG); or not brought to trial or *nolle prosequi* (NP). In the analysis, we shall consider how this variable is affected by factors corresponding to type of offense for which the defendant was charged, the county where the case was tried, and the race of the defendant.[14] The categories associated with these factors are as follows:

Type of Offense	*County*	*Race*
Public Drunkenness (A)	Durham (D)	Black (B)
Violence against Persons (V)	Orange (O)	White (W)
Property (P)		
Major Traffic (MT)		
Speeding (S)		

Two functions which are of interest for describing the outcome are: (1) the relative number of cases that are not brought to trial (NP) to those that are (G or NG), and (2) the relative number of cases where the defendant pleads guilty (G) to those where the defendant pleads not guilty (NG). Instead of dealing with these ratios directly, we shall work with their corresponding logarithms, which have some desirable mathematical properties. First, the ratio $R = (F_1/F_2)$ is a nonlinear function of F_1 and F_2 but

[11] The Governor's Committee on Law and Order, *The Distribution, Detection, and Disposition of Criminal Offenses in North Carolina,* (Department of Local Affairs, State of North Carolina, 1969).

[12] R.G. Lehnen and J.O. Williams, "Some Aspects of the Criminal Court Process in North Carolina," *North Carolina Law Review,* 49 (April 1971), 469–85.

[13] For Durham County, the sample covers the time period from January 1, 1967, to April 30, 1969; for Orange County, the sample covers the period from January 1, 1969, to April 30, 1969.

[14] The District Courts are organized along county lines.

TABLE 8

Pleas to Criminal Charges for County X Race X Offense Subpopulations

Case Situation				*Defendant's Plea*			*Preliminary Estimated Proportion*		
Offense	County	Race	Total	G	NG	NP	G	NG	NP
A	D	B	45	33	8	4	.733	.178	.089
A	D	W	57	53	2	2	.930	.035	.035
A	O	B	16	5	10	1	.313	.625	.062
A	O	W	16	14	2	0	.857	.122	.020
V	D	B	23	10	10	3	.435	.435	.130
V	D	W	16	7	8	1	.438	.500	.062
V	O	B	15	5	5	5	.333	.333	.333
V	O	W	13	1	5	7	.077	.385	.538
P	D	B	19	9	8	2	.474	.421	.105
P	D	W	17	10	5	2	.588	.294	.118
P	O	B	19	11	5	3	.579	.263	.158
P	O	W	9	5	4	0	.536	.429	.035
MT	D	B	7	4	2	1	.571	.286	.143
MT	D	W	21	16	3	2	.762	.143	.095
MT	O	B	19	12	6	1	.632	.316	.053
MT	O	W	27	13	13	1	.481	.481	.037
S	D	B	35	32	3	0	.906	.085	.009
S	D	W	95	87	5	3	.916	.053	.032
S	O	B	25	20	3	2	.800	.120	.080
S	O	W	121	98	16	7	.810	.132	.058

Offense Situation	Counties	Race	Type of Plea
A: Public Drunkenness	D: Durham	B: Black	G: Guilty
V: Violence	O: Orange	W: White	NG: Not Guilty
P: Property Related			NP: *Nolle Prosequi*
MT: Major Traffic			
S: Speeding			

$\log_e(R) = \log_e(F_1) - \log_e(F_2)$ is a linear function of $\log_e(F_1)$ and $\log_e(F_2)$. Second, $\log_e(R)$ reflects the relationship of F_1 to F_2 in a more symmetric manner, since $\log_e(R)$ varies from $-\infty$ to 0 when F_2 exceeds F_1 and varies from 0 to $+\infty$ when F_1 exceeds F_2, whereas R varies from 0 to +1 when F_2 exceeds F_1 and varies from +1 to $+\infty$ when F_1 exceeds F_2.

For each Offense X County X Race subpopulation, the log-ratio of NP vs.

(G or NG) and the log-ratio of G vs. NG can be formulated in terms of Equations 1.6 and 1.7 with the **A** and **K** matrices in Equations 2.5 and 2.6. These estimates

$$\underset{80\times 60}{\mathbf{A}} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_1 & \dots & \mathbf{0} \\ \dots & \dots & \dots & \dots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{A}_1 \\ \mathbf{A}_2 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_2 & \dots & \mathbf{0} \\ \dots & \dots & \dots & \dots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{A}_2 \end{bmatrix} \quad \text{where} \quad \begin{aligned} \underset{2\times 3}{\mathbf{A}_1} &= \begin{bmatrix} 0 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} \\ \underset{2\times 3}{\mathbf{A}_2} &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \\ \underset{2\times 3}{\mathbf{0}} &= \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \end{aligned} \tag{2.5}$$

and

$$\underset{40\times 80}{\mathbf{K}} = \begin{bmatrix} \mathbf{K}_1 & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{K}_1 & \dots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{K}_1 & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{K}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} & \mathbf{K}_1 & \dots & \mathbf{0} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} & \mathbf{0} & .. & \mathbf{K}_1 \end{bmatrix} \quad \text{where} \quad \begin{aligned} \underset{(1\times 2)}{\mathbf{K}_1} &= [1 \ -1] \\ \underset{(1\times 2)}{\mathbf{0}} &= [0 \ 0] \end{aligned} \tag{2.6}$$

are shown in Table 9 under the heading "Preliminary Estimates." The corresponding estimated covariance matrix is obtained by applying Eq. A.2; for the situation where **A** and **K** are defined as in Eq. 2.5 and 2.6, it has a diagonal structure—*i.e.*, all the nondiagonal elements are zero. The respective diagonal elements, which represent the set of estimated variances, are shown in Table 9. Finally, it can be noted that in obtaining all of these estimates the 0-frequencies in Table 8 have been replaced by (1/3) since the logarithm of 0 is not defined.

It was anticipated that Type of Offense would be the dominant factor affecting both $\log_e\{NP/(G + NG)\}$ and $\log_e\{G/NG\}$, and there would be potential interactions with both County and Race. For this reason, the first model which was applied to these estimated functions was the "incomplete hierarchical" model specified by the matrix $\mathbf{X}_1$ in Eq. 2.7. The rows of $\mathbf{X}_1$ may be

TABLE 9
Estimates For Hierarchical Court Model

			Preliminary Estimates		*Variance Estimates*		*Predicted Probabilities from $\mathbf{X}_2$ Model*		
Offense	County	Race	$\log_e\left\{\frac{NP}{G+NG}\right\}$	$\log_e\left\{\frac{G}{NG}\right\}$	$\log_e\left\{\frac{NP}{G+NG}\right\}$	$\log_e\left\{\frac{G}{NG}\right\}$	Guilty	Not Guilty	Nolle Pros.
A	D	B	−2.33	1.42	.2744	.1553	.7453	.2010	.0537
A	D	W	−3.31	3.28	.5182	.5189	.9264	.0199	.0537
A	O	B	−2.71	−0.69	1.0667	.3000	.5413	.4050	.0537
A	O	W	−3.88	1.95	3.0928	.5714	.8931	.0532	.0537
V	D	B	−1.90	0.00	.3833	.2000	.4126	.4793	.1081
V	D	W	−2.71	−0.13	1.0667	.2679	.4126	.4793	.1081
V	O	B	−0.69	0.00	.3000	.4000	.2635	.3060	.4305
V	O	W	0.15	−1.61	.3095	1.2000	.2635	.3060	.4305
P	D	B	−2.14	0.12	.5588	.2361	.5874	.3128	.0998
P	D	W	−2.01	0.69	.5667	.3000	.5874	.3128	.0998
P	O	B	−1.67	0.79	.3958	.2909	.5874	.3128	.0998
P	O	W	−3.31	0.22	3.1414	.4500	.5874	.3128	.0998
MT	D	B	−1.79	0.69	1.1667	.7500	.6821	.2181	.0998
MT	D	W	−2.25	1.67	.5526	.3958	.6821	.2181	.0998
MT	O	B	−2.89	0.69	1.0556	.2500	.4771	.4231	.0998
MT	O	W	−3.26	0.00	1.0385	.1539	.4771	.4231	.0998
S	D	B	−4.66	2.37	3.0589	.3646	.8790	.0673	.0537
S	D	W	−3.42	2.86	.3442	.2115	.8790	.0673	.0537
S	O	B	−2.44	1.90	.5435	.3833	.7820	.1643	.0537
S	O	W	−2.79	1.81	.1516	.0727	.7820	.1643	.0537

$$\underset{40\times30}{\mathbf{X}_1} = \begin{bmatrix} \mathbf{X}_{11} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_{11} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{X}_{11} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{X}_{11} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{X}_{11} & \cdots & \mathbf{0} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{X}_{11} \end{bmatrix} \quad \underset{4\times3}{\mathbf{X}_{11}} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & -1 & 1 \\ 1 & -1 & -1 \end{bmatrix} \quad \underset{4\times3}{\mathbf{0}} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{2.7}$$

identified with the respective subpopulations, first for the $\log_e \{NP/(G+NG)\}$ function and then for the $\log_e \{G/NG\}$ functions. The columns of $\mathbf{X}_1$ correspond to the sources of variance in Table 10 where statistical tests of significance are also given. In Table 10 the Residual X^2 statistics are nonsignificant ($\alpha = .25$), which implies that this model provides a satisfactory

TABLE 10

Analysis of Variance for Hierarchical Court Model

Source of Variation	Single Function D.F.	Single Function X^2: $\log_e \left\{\frac{NP}{G+NG}\right\}$	Single Function X^2: $\log_e \left\{\frac{G}{NG}\right\}$	Overall D.F.	Overall X^2
Type of Offense	4	17.61	44.33	8	61.94
County for A	1	.19	11.01	2	11.20
County for V	1	6.74	.43	2	7.17
County for P	1	.01	.13	2	.14
County for MT	1	1.18	3.29	2	4.47
County for S	1	1.56	3.81	2	5.37
Race for A	1	1.55	12.73	2	14.28
Race for V	1	.30	.60	2	.90
Race for P	1	.10	.03	2	.13
Race for MT	1	.19	.22	2	.41
Race for S	1	.01	.11	2	.12
Overall Model	14	38.50	81.44	28	119.94
Residual (Combined County × Race)	5	2.62	4.59	10	7.21
Total	19	41.12	86.03	38	127.15

TABLE 11

Estimates of Effects for Hierarchical Court Model

Offense Situation	$\log_e\left\{\frac{NP}{G+NG}\right\}$ Offense	County	Race	$\log_e\left\{\frac{G}{NG}\right\}$ Offense	County	Race
A	−3.03	0.22	0.51	1.50	0.94	−1.10
V	−1.15	−0.88	−0.18	−.31	0.21	0.23
P	−2.02	−0.05	0.15	.47	−0.10	−0.05
MT	−2.55	0.52	0.21	.83	0.54	0.13
S	−3.10	−0.41	0.04	2.21	0.44	−0.08

Offense effects are mean effects where the specified offense applies.
County effect is added for Durham County and subtracted for Orange County.
Race effect is added for blacks and subtracted for whites.

characterization of the data. This result may also be interpreted to mean that the County × Race interaction and the Offense × County × Race interaction are negligible. Otherwise, it is apparent that the effects of County and Race are more difficult to evaluate. For this purpose, some insights can be gained from the estimated effects in Table 11. Hence, for the $\log_e\{NP/(G + NG)\}$ function, the only significant ($\alpha = .05$) County effect is that corresponding to "Violence against Persons"; and none of the race effects are significant ($\alpha = .05$). In addition, the offense effects for "Public Drunkenness" and "Speeding" are similar, as well as those for "Property" and "Major Traffic."

For the $\log_e\{G/NG\}$ function, the only significant County effect ($\alpha = .01$) is that corresponding to "Public Drunkenness"; however, the estimated county effect for this category is similar to those for "Major Traffic" and "Speeding," which are nearly significant ($\alpha = .05$). Otherwise, the only significant Race effect is that corresponding to "Public Drunkenness," and the Offense effects may be classified into the same pattern that applied to $\log_e\{NP/(G+NG)\}$. All of these conclusions should be regarded cautiously because some of the apparently nonsignificant effects could in future studies with larger sample sizes be demonstrated to be important. Nevertheless, the results which have been obtained suggest that the important effects with respect to the estimated functions in this example can be summarized in terms of the model $\mathbf{X}_2$ in Eq. 2.8, with $\mathbf{0}_{21}$ and $\mathbf{0}_{22}$ being (20×4) and (20×5) matrices of 0's respectively.

$$\underset{(40\times 9)}{\mathbf{X}_2} = \begin{bmatrix} \mathbf{X}_2 & \mathbf{0}_{21} \\ \mathbf{0}_{22} & \mathbf{X}_{22} \end{bmatrix} \text{ where } \underset{(20\times 4)}{\mathbf{X}_{21}} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & -1 & -1 \\ 1 & -1 & -1 & -1 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{bmatrix} \text{ and } \underset{(20\times 5)}{\mathbf{X}_{22}} = \begin{bmatrix} 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & -1 \\ 1 & 1 & 0 & -1 & 1 \\ 1 & 1 & 0 & -1 & -1 \\ 1 & -1 & -1 & 0 & 0 \\ 1 & -1 & -1 & 0 & 0 \\ 1 & -1 & -1 & 0 & 0 \\ 1 & -1 & -1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & -1 & 0 \\ 1 & 0 & 1 & -1 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & -1 & 0 \\ 1 & 1 & 0 & -1 & 0 \end{bmatrix} \quad (2.8)$$

The effects in this model which have a nonsignificant Residual X^2 statistic ($\alpha = .05$), together with appropriate statistical tests, are indicated in Table 12. Moreover, the ratio of the Overall Model X^2 statistic vs. the sum of those for the Overall Model and Residual is $105.69/(105.69 + 21.46) = 0.83$, which may be interpreted to mean that the model (2.8) accounts for 83% of the total weighted variation (in the sense of Q) among the observed cell frequencies, as reflected by the $\log_e\{NP/(G + NG)\}$ and $\log_e\{G/NG\}$ functions.

The final court model shows that the type of offense was the principal source of variation in the rates of trial [NP vs. (G + NG)]. Persons charged with Public Drunkenness or Speeding were most likely to have their cases tried, followed next by Property and Major Traffic, and last by Personal Violence. The race of the defendant had no effect on the likelihood of trial, and except in one instance, the location of the District Court also made no difference. The exception was for crimes of Personal Violence, where Durham County was more likely to bring such cases to trial than was Orange County.

For those defendants who were tried and entered pleas (G vs. NG), the

TABLE 12

Analysis of Variance for Reduced Court Model

Source of Variation	D.F.	X^2
Model for $\log_e \left\{\frac{NP}{G+NG}\right\}$	3	32.53
i. Type of Offense $\{$(A or S) vs. V vs. (P or MT)$\}$	2	17.30
ii. County for V	1	7.76
Residual for $\log_e \left\{\frac{NP}{G+NG}\right\}$	16	8.59
Model for $\log_e \left\{\frac{G}{NG}\right\}$	4	73.16
i. Type of Offense (A or S) vs. V vs. (P or MT)	2	56.68
ii. County for (A or S or MT)	1	12.45
iii. Race for A	1	19.37
Residual for $\log_e \left\{\frac{G}{NG}\right\}$	15	12.87
Combined Overall Model	7	105.69
Combined Overall Residual	31	21.46

offense charged made the most difference whether or not a guilty plea was entered. Public Drunkenness and Speeding cases had the highest rate of guilty pleas, followed next by Property and Major Traffic cases, and last by Personal Violence cases. Race was a factor affecting pleas on charges relating to Public Drunkenness, blacks pleading guilty less often than whites. Defendants tried in Durham County were more likely to plead guilty in Public Drunkenness, Speeding, and Major Traffic cases than those tried in Orange County.

Finally, predicted values for $\log_e\{NP/(G + NG)\}$ and $\log_e\{G/NG\}$ can be determined by applying Eq. A.7 to the estimated parameters for the model Eq. 2.8. These results can then be transformed to predicted cell probabilities by noting that

$$\begin{aligned} 1 &= NP + G + NG, \\ \lambda &= \log_e\{NP/(G + NG)\}, \text{ and} \\ \phi &= \log_e\{G/NG\} \end{aligned}$$

implies that

$$\begin{aligned} NP &= e^{\lambda}/(1 + e^{\lambda}) \\ G &= e^{\phi}/(1 + e^{\phi})(1 + e^{\lambda}), \text{ and} \\ NG &= 1/(1 + e^{\phi})(1 + e^{\lambda}) \end{aligned}$$

The predicted probabilities obtained in this manner are shown in Table 9 and represent refined estimates of the preliminary estimated proportions given in

Table 8 in accordance with the model Eq. 2.8. Thus the methodology we have presented has permitted a reasonably detailed analysis to be directed at the factors affecting an aspect of behavior in the criminal court process.

Some Concluding Comments

The method of statistical analysis for complex contingency tables presented in this paper is general and widely applicable. It requires essentially no assumptions other than one requiring that the table be generated by either random sampling with known probabilities of selection or from a product-multinomial distribution. Furthermore, the statistical procedures and tests of significance used are approximate only in the sense of being based on large sample theory. Many statistical procedures familiar to social scientists, such as the Student's t test for the difference of means, are based on similar foundations.

Some of the traditional parametric methods share several of the advantages of the GSK method. Analysis of variance of qualitative data, while having some heuristic value, provides descriptive statistics but no valid tests of significance. The validity of the tests of significance is compromised in such applications of analysis of variance, because the hypotheses of interest do not have a suitable estimate of the variance matrix **V**. Since the GSK approach is directed at the estimated proportions based on large sample sizes, an acceptable variance matrix is available for providing tests of significance and confidence intervals.

In some situations where the dependent variable(s) are nominal, few suitable techniques exist that permit a comprehensive analysis of the data. Notwithstanding the problem mentioned above, analysis of variance and dummy variable regression, if not inappropriate to apply, generally have ambiguous results.

Other advantages in addition to the estimates, standard errors and confidence intervals, and residuals provided by the general linear approach are worth noting. The GSK approach offers great flexibility in the analysis of problems with varying structures and constraints. Furthermore, the matrix input approach used in the computer program is an excellent teaching device for illustrating similarities and differences in analyses, for it highlights conceptual problems made unclear by the other approaches. Thus, one may consider whether a complex contingency table presents a "multifactor, single response" problem or a "multiresponse" problem and the implications of each alternative.

The goodness-of-fit test provides another distinct advantage. Although most conventional techniques do not provide a very explicit criterion for including and excluding parameters for study, the Chi-square goodness-of-fit test associated with each model fitted to the complex contingency table provides an explicit standard for fitting a new model or working with the one at hand. Such a model-fitting approach is relatively unknown to social scientists, who rarely consider how well the estimated parameters provide an adequate prediction of the observed scores. One of the more familiar examples is associated with fitting regressions, where some arbitrarily defined level of "explained variance" is used to justify the conclusion that the model is doing an adequate job of accounting for the variation of the dependent variable, when in fact, it may not.

Although political scientists have revolutionized their approaches to the study of politics, in many respects they continue to use techniques that are less and less suited to their increasingly complex needs. The unique combination of advantages associated with the general linear approach to nonmetric data make it a more precise method for meeting these analytical needs and, where suitable data exist, for confirming and building social theories.

APPENDIX

The variance-covariance matrix $\mathbf{V}$ (rs $\times$ rs) of the cell probabilities has been defined previously in expressions (1.1–1.4). If the functions of interest are linear as described in (1.6), the covariance matrix can be consistently estimated by

$$\mathbf{V_a} = \mathbf{AVA'} \tag{A.1}$$

For the log-linear functions described in (1.7), the corresponding estimated covariance matrix is

$$\mathbf{V_f} = \mathbf{KD_a^{-1}\ AVA'D_a^{-1}K'} \tag{A.2}$$

where $\mathbf{D_a}$ is a diagonal matrix with elements of the vector $\mathbf{a}$ on the main diagonal. Thus, $\mathbf{V_F}$ is equal to $\mathbf{V_a}$ when linear functions are used and equal to $\mathbf{V_f}$ whenever log-linear functions are used.

When fitting a linear model as described in (1.11) by means of weighted least squares, the estimated **b** are obtained by minimizing the quadratic function[15]

$$Q = (\mathbf{F} - \mathbf{Xb})' \mathbf{V_F^{-1}} (\mathbf{F} - \mathbf{Xb}) \tag{A.3}$$

and belong to the class of best asymptotic normal (BAN) estimators as discussed by Neyman.[16] The resulting expression for **b** is

$$\mathbf{b} = (\mathbf{X}' \mathbf{V_F^{-1}} \mathbf{X})^{-1} \mathbf{X}' \mathbf{V_F^{-1}} \mathbf{F} \tag{A.4}$$

A consistent estimate for the covariance matrix of **b** is given by

$$\mathbf{V_b} = (\mathbf{X}' \mathbf{V_F}^{-1} \mathbf{X})^{-1} \tag{A.5}$$

A goodness-of-fit statistic for assessing the extent to which (1.11) characterizes **F** is the residual sum of squares in (A.3) with **b** replaced by the estimate in (A.4). Under the hypothesis that the model fits, Q has approximately a chi-square distribution with D.F. = (g – t) in large samples.

If the model does adequately describe the data, tests of linear hypotheses involving **b** can be undertaken. In particular, for a general hypothesis of the form $H_0 : \mathbf{Cb} \hat{=} 0$ where **C** is a known (d × t) coefficient matrix of full rank $d \leq t$, a suitable test statistic is

$$X^2 = SS\,(\mathbf{Cb} \hat{=} \mathbf{0}) = \mathbf{b}'\, \mathbf{C}'\, [\mathbf{C}\, (\mathbf{X}' \mathbf{V_F^{-1}} \mathbf{X})^{-1} \mathbf{C}']^{-1}\, \mathbf{C\, b} \tag{A.6}$$

which has approximately a chi-square distribution with D. F. = d in large samples under H_0.

Also, when a model like (1.11) fits, it is useful to determine predicted values $\hat{\mathbf{F}}$ for **F** by using the expression

$$\hat{\mathbf{F}} = \mathbf{X}\, (\mathbf{X}' \mathbf{V_F^{-1}} \mathbf{X})^{-1} \mathbf{X}' \mathbf{V_F^{-1}} \mathbf{F} = \mathbf{Xb} \tag{A.7}$$

[15] It should be noted that the functions **F** must be determined in a manner so that $\mathbf{V_F}$ is nonsingular. In this regard, some guidance as to the proper choice of **A** and **K** will be given in the examples in Section 3. Otherwise, see Grizzle, Starmer, and Koch, "Analysis of Categorical Data by Linear Models," 489–504.

[16] J. Neyman, "Contributions to the Theory of the χ^2 Test," *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability,* (Berkeley and Los Angeles: University of California Press, 1949), 239–72.

The functions $\hat{\mathbf{F}}$ represent improved estimates of the same parameters which are originally estimated by **F** since they have been efficiently determined from an analysis which simultaneously encompasses all of the available data. An estimate for the covariance matrix of $\hat{\mathbf{F}}$ is given by

$$\mathbf{V}_{\hat{\mathbf{F}}} = \mathbf{X}\,[\mathbf{X}'\,\mathbf{V}_{\mathbf{F}}^{-1}\,\mathbf{X}]^{-1}\,\mathbf{X}' \qquad (A.8)$$

and the extent to which the diagonal elements of $\mathbf{V}_{\hat{\mathbf{F}}}$ are uniformly smaller than the diagonal elements of $\mathbf{V}_{\mathbf{F}}$ reflects the gain in precision regarding knowledge of **F** which is due to the use of the linear model (1.11).

Finally, in certain types of applications, the occurrence of situations where some of the n_{ij} are zero may cause difficulty by either inducing singularities in $\mathbf{V}_{\mathbf{F}}$ or by preventing the calculation of certain log-linear functions. One way to deal with this is to replace these zero frequencies by $(1/r)$. Other aspects of this problem are discussed by Berkson[17] and GSK.

All of the calculations which are used can be formulated in terms of standard matrix multiplication and matrix inversion operations and thus can be readily performed.[18] With respect to theoretical properties, the goodness-of-fit statistic based on (1.10) and the tests involving estimated parameters in (1.11) belong to the class of minimum modified chi-square statistics due to Neyman[19] which is equivalent to the general quadratic form criteria of Wald.[20] Two alternative approaches

[17] Joseph Berkson, "Approximation of Chi-Square by 'Probits' and by 'Logits'," *Journal of the American Statistical Association,* 41 (March 1946), 70–74; and Berkson, "Maximum Likelihood and Minimum χ^2 Estimate of the Logistic Function," *Journal of the American Statistical Association,* 50 (March 1955), 130–162.

[18] Computer programs permitting the efficient calculation of the estimates and test statistics discussed here are not difficult to prepare. The one used for the matrix operations in the analyses of this paper can be obtained from the Program Librarian, Department of Biostatistics, School of Public Health, University of North Carolina, Chapel Hill, North Carolina 27514. For a program description see R.N. Forthofer, C.F. Starmer, and J.E. Grizzle, "A Program for the Analysis of Categorical Data by Linear Models," *University of North Carolina Institute of Statistics Memo Series,* No. 604 (January 1969), or a published description by the same authors in *Journal of Biomedical Systems,* 2 (Spring 1971), 3–48.

[19] Neyman, "Contributions to the Theory of the χ^2 Test," 239–72.

[20] A. Wald, "Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations Is Large," *Transactions American Mathematical Society,* 54 (July to December 1943), 426–82.

to this methodology are that based on maximum likelihood as formulated by Bishop[21] and Goodman[22] and that based on minimum discrimination information as formulated by Ku, Varner, and Kullback.[23] These two procedures also have a regression flavor to them, but calculations are in terms of multiplicative and possibly iterative operations on appropriate marginal distributions. In large samples, all three of these methods are asymptotically equivalent in the sense of being based on BAN estimates. Thus, in most applications the choice of which method to use is a matter of practical and computational convenience.

Manuscript submitted May 8, 1972.
Final manuscript received October 10, 1972.

[21] Yvonne M.M. Bishop, "Full Contingency Tables, Logits, and Split Contingency Tables," *Biometrics,* 25 (June 1969), 383–99; and Bishop, "Effects of Collapsing Multidimensional Contingency Tables," *Biometrics,* 27 (September 1971), 545–62.

[22] L.A. Goodman, "The Multivariate Analysis of Qualitative Data: Interactions among Multiple Classifications," *Journal of the American Statistical Association,* 65 (March 1970), 226–56; Goodman, "The Analysis of Multidimensional Contingency Tables: Stepwise Procedures and Direct Estimation Methods for Building Models for Multiple Classifications," *Technometrics,* 13 (February 1971), 33–61; Goodman, "Partitioning of Chi-Square, Analysis of Marginal Contingency Tables, and Estimation of Expected Frequencies in Multidimensional Contingency Tables," *Journal of the American Statistical Association,* 66 (June 1971), 339–44; Goodman, "A Modified Multiple Regression Approach to the Analysis of Dichotomous Variables," *American Sociological Review,* 37 (February 1972), 28–46; and Goodman, "A General Model for the Analysis of Surveys," *American Journal of Sociology,* 77 (May 1972), 1035–1086.

[23] H.H. Ku, R. Varner, and S. Kullback, "Analysis of Multidimensional Contingency Tables," *Journal of the American Statistical Association,* 66 (March 1971), 55–64.