This research was supported by the National Institutes of Health, Institute of General Medical Sciences Grants GM-70004-01, GM-0038-18, and GM-12868-08.

LINEAR MODEL ANALYSIS OF CATEGORICAL DATA WITH INCOMPLETE RESPONSE VECTORS

by

Gary G. Koch, 1 Peter B. Imrey, 1 and Donald W. Reinfurt 2

1Department of Biostatistics and 2Highway Safety Research Center
University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series No. 790

December 1971
SUMMARY

The general linear model approach to the analysis of categorical data described by Grizzle, Starmer and Koch [1969] is extended to situations where: (1) missing data for certain individuals arise at random as a result of non-response or deleted incorrect response; (2) supplemental samples pertaining to various subsets of variables have been obtained due to cost considerations and/or special interest in these variables. The problems discussed are distinct from those involving "incomplete contingency tables" containing a priori empty cells.

The extension is presented through a series of examples which show how the approach can be used to handle a wide variety of non-standard data configurations. Applications to categorical data mixed models and split plot designs are emphasized.
LINEAR MODEL ANALYSIS OF CATEGORICAL DATA WITH INCOMPLETE RESPONSE VECTORS

Gary G. Koch\textsuperscript{1}, Peter B. Imrey\textsuperscript{2}, and Donald W. Reinfurt\textsuperscript{2}
Department of Biostatistics and Highway Safety Research Center
University of North Carolina, Chapel Hill, N.C. 27514, U.S.A.

1. INTRODUCTION

In a recent paper, Grizzle, Starmer, and Koch [1969] (subsequently abbreviated GSK) described how linear regression models and weighted least squares could be used to analyze multivariate categorical data which has been summarized in a multi-dimensional contingency table. This methodology provides the researcher with a powerful and flexible tool which can be used to evaluate models and to test hypotheses for a broad class of experimental situations. For example, various conditions of "no interaction" as discussed by Roy and Kastenbaum [1956] and Bhapkar and Koch [1968a, 1968b] can be investigated either as they apply to cell probabilities or to certain functions thereof (e.g., marginal probabilities, logits, mean scores, etc.). The resulting test statistics belong to the class of minimum modified $- \chi_1^2$, due to Neyman [1949], and are equivalent to certain generalized quadratic form criteria of Wald [1943]. Alternatively, many of these same problems can be approached by using the methods described by Mantel [1966], Lewis [1968], Bishop [1969, 1971], Fienberg [1970], and Goodman [1970, 1971a, 1971b] based on maximum likelihood, or those described by Ku, Varner, and Kullback [1968, 1971] based on minimum discrimination information.

All of the previously mentioned papers are primarily concerned with data where there is complete classification of experimental units (or subjects) with respect to all variables of interest. However, in certain types of applications incomplete configurations are encountered either due to chance or design. Examples here include:
1. Cases where missing data arise as a result of non-response or deleted incorrect response, but such missing data can be presumed to occur at random.

2. Cases where supplemental samples pertaining to a subset of variables have been obtained either because of greater interest in those variables or economic cost considerations. These situations are analogous to those which arise in the context of augmented fractional factorial designs as discussed by Box and Wilson [1951], Box [1966], John [1966], Gaylor and Merrill [1968], and sequences of fractional factorials as discussed by Daniel [1962] and Addelman [1969]; this relationship is particularly apparent when one is concerned with a categorical data mixed model or split plot experiment as described by Koch and Reinfurt [1970].

3. Incomplete block or fractional factorial type split plot multivariate designs like those described by Roy, Gnanadesikan, and Srivastava [1971].

4. Growth curve experiments like those described by Potthoff and Roy [1964] and Allen and Grizzle [1969].

The principal objective of the analysis for any of these situations is to obtain valid and precise estimators for all of the relevant parameters in a manner which uses as much of the available information in the data as possible (i.e., subjects with incomplete responses are not necessarily omitted).

With respect to incomplete categorical data, Hocking and Oxspring [1971] have discussed maximum likelihood for the case of sampling from a single multinomial population in terms of an approach similar to that given by Hocking and Smith [1968] for incomplete multivariate normal data. A similar problem has been considered by Blumenthal [1968], also in terms of maximum likelihood. Alternatively, Koch and Reinfurt [1970] have described a two-stage procedure which yields the
minimum-$\chi^2$ analysis of incomplete categorical data. The remainder of this paper will be concerned with describing some additional aspects of this extension of the GSK approach and with illustrating its application to several pertinent examples.

Finally, it is important to note that the situations under consideration here are entirely different from those corresponding to "incomplete contingency tables" as discussed by Goodman [1968], Bishop and Fienberg [1969], Williams and Grizzle [1970], and Mantel [1970]. In these cases, each experimental unit is classified according to all of the variables of interest, but certain combinations of such variables have zero probability of occurrence. Thus, the resulting contingency table is incomplete in the sense that certain cells contain zero frequencies regardless of sample size. Hence, the term "incomplete contingency table" is a suitable descriptive title for such situations.

However, the question then arises as to what to call the incomplete categorical data situations of interest in the remainder of this paper. For lack of a better term cases like those cited in (1) or (2) will be referred to as "augmented contingency tables" while the ones like those cited in (3) or (4) will be called "complex split plot contingency tables" depending on the nature of the experimental situation. Some of the reasoning behind these descriptive titles will be apparent from the nature of the examples which will be considered.

2. SOME RELEVANT EXAMPLES

2.1 An application with missing data (non-response)

In Table 1, data are given for the unaided distance vision of 8577 women aged 30-39. For 7477 of these women, vision grade was classified for both eyes and the resulting frequencies are identical to the data used as an illustrative example for tests of marginal homogeneity by several authors.
TABLE 1

UNAIDED DISTANCE VISION; 8577 WOMEN AGED 30-39

<table>
<thead>
<tr>
<th>Right Eye</th>
<th>Highest Grade (1)</th>
<th>Second Grade (2)</th>
<th>Third Grade (3)</th>
<th>Lowest Grade (4)</th>
<th>Sub-Total</th>
<th>Right Only</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Highest grade (1)</td>
<td>1520</td>
<td>266</td>
<td>124</td>
<td>66</td>
<td>1976</td>
<td>140</td>
<td>2116</td>
</tr>
<tr>
<td>Second grade (2)</td>
<td>234</td>
<td>1512</td>
<td>432</td>
<td>78</td>
<td>2256</td>
<td>150</td>
<td>2406</td>
</tr>
<tr>
<td>Third grade (3)</td>
<td>117</td>
<td>362</td>
<td>1772</td>
<td>205</td>
<td>2456</td>
<td>160</td>
<td>2616</td>
</tr>
<tr>
<td>Lowest grade (4)</td>
<td>36</td>
<td>82</td>
<td>179</td>
<td>492</td>
<td>789</td>
<td>50</td>
<td>839</td>
</tr>
<tr>
<td>Sub-Total</td>
<td>1907</td>
<td>2222</td>
<td>2507</td>
<td>841</td>
<td>7477</td>
<td>*</td>
<td>7977</td>
</tr>
<tr>
<td>Left Only</td>
<td>160</td>
<td>180</td>
<td>200</td>
<td>60</td>
<td>600</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Total</td>
<td>2067</td>
<td>2402</td>
<td>2707</td>
<td>901</td>
<td>8077</td>
<td>*</td>
<td>8577</td>
</tr>
</tbody>
</table>

including Stuart [1955], Bhapkar [1966], Ireland, Ku and Kullback [1969] and GSK.

The remaining 1100 observations form artificial data for 600 women for whom only left eye vision was reported and 500 women for whom only right eye vision was reported. It will be presumed that the incomplete data for women with vision classified only for one eye arose in a completely random manner which was statistically independent of the true classification of their vision with respect to both eyes. This assumption allows us to say that the marginal probabilities pertaining to left eye vision and right eye vision for women classified on both eyes are the same parameters as the probabilities pertaining to left eye vision for women classified only for the left eye and to right eye vision for women classified only for the right eye respectively.

To apply the GSK approach to this problem, we view the data as coming from three populations and arrange it in an array as follows:
Let $\pi_{jj'}$ denote the probability that a subject in the population is classified according to the $j$-th category on the right eye and the $j'$-th category on the left eye where $j, j' = 1, 2, 3, 4$; and let $\pi_j$ and $\pi_j'$ represent corresponding marginal probabilities with

$$\pi_j = \sum_{j' = 1}^{4} \pi_{jj'} \quad \text{and} \quad \pi_j' = \sum_{j = 1}^{4} \pi_{jj'}$$

If we define $p = (n/n)$, $p_R = (n_R/n)$, and $p_L = (n_L/n)$ where $n = 7477$, $n_R = 500$, and $n_L = 600$ are the sums of the elements in $\pi$, $\pi_R$, $\pi_L$ respectively and if we let $p_G' = (p', p_R', p_L')$ be the augmented vector of observed relative frequencies, then

$$E\{p_G\} = E\begin{bmatrix} \pi_p \\ \pi_{p_R} \\ \pi_{p_L} \end{bmatrix} = \begin{bmatrix} \pi' \\ \pi_{1*}' \\ \pi_{2*}' \end{bmatrix}$$

(2.1)

where

$$\pi' = (\pi_1, \pi_2, \pi_3, \pi_4, \pi_1, \pi_2, \pi_3, \pi_4, \pi_1, \pi_2, \pi_3, \pi_4)$$

$$\pi_{1*}' = (\pi_1, \pi_2, \pi_3, \pi_4)$$

$$\pi_{2*}' = (\pi_1, \pi_2, \pi_3, \pi_4)$$

In addition, the covariance matrix for $p_G$ is given by

$$\Sigma_{\pi} = \text{Var}_{\pi} \begin{bmatrix} p_p \\ p_{p_R} \\ p_{p_L} \end{bmatrix} = \begin{bmatrix} (D_{\pi} - \pi_p)'/n \\ 0_{4, 16} \\ 0_{16, 4} \\ (D_{\pi} - \pi_p)'/n \\ 0_{4, 16} \\ 0_{16, 4} \\ (D_{\pi} - \pi_{1*}')/n \\ 0_{4, 16} \\ 0_{16, 4} \end{bmatrix}$$

(2.2)

where $0_{k, k}$ denotes a $(k \times k)$ zero matrix and $D_{\pi}$, $D_{\pi_R}$, $D_{\pi_L}$ are diagonal matrices with diagonal elements being corresponding elements of $\pi$, $\pi_{1*}$, $\pi_{2*}$. The matrix $\Sigma_{\pi}$ can be consistently estimated by replacing $\pi, \pi_{1*}, \pi_{2*}$ with $p, p_R, p_L$ respectively in (2.2); i.e., forming $\Sigma_{\pi_G}$.
The estimate $\hat{p} = (n/n)$ of $\tau$ in (2.1) can be adjusted to reflect the available data for all 8577 women by applying weighted least square to the vector $p_G$ with weights derived from the matrix $V(p_G)$ in the manner described by GSK. In order to avoid computational singularities which arise because the sum of the elements in each of the vectors $p$, $p_R$, $p_L$ is unity, the analysis is applied to the vector $\sim p_G = \sim A p_G$ where

$$A = \begin{bmatrix}
I_{15} & 0_{15,3} & 0_{15,3} & 0_{15,3} & 0_{15,3} \\
0_{3,15} & I_{3} & 0_{3,1} & 0_{3,1} & 0_{3,1} \\
0_{3,15} & 0_{3,1} & I_{3} & 0_{3,1} & 0_{3,1} \\
0_{3,15} & 0_{3,1} & 0_{3,1} & I_{3} & 0_{3,1} \\
0_{3,15} & 0_{3,1} & 0_{3,1} & 0_{3,1} & I_{3}
\end{bmatrix}$$

with $I_{3}$ and $I_{15}$ being identity matrices and the other entries being suitable zero matrices. For the vector $\sim F$, the corresponding weights are derived from $\sim V_F = A[\sim V(p_G)]A^t$. From (2.1), it follows that the linear model which pertains to $\sim F$ is given by $\sim E(\sim F) = \sim X\beta$ where $\sim X$ is the known coefficient

$$E(\sim F) = \begin{bmatrix}
I_{15} \\
1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix} \beta, \quad (2.3)$$

matrix specified in (2.3) and $\beta$ is the vector of the first 15 components of $\tau$; i.e. $\tau$ with $\tau_{44}$ deleted. The resulting estimate $\sim b$ of $\beta$ is given by

$$b = (\sim X'\sim V_F^{-1}\sim X)^{-1} \sim X'\sim V_F^{-1} \sim F$$

with estimated covariance matrix

$$\sim V_b = \text{Var}(\sim b) = (\sim X'\sim V_F^{-1}\sim X)^{-1} \quad (2.5)$$
A goodness of fit statistic for assessing the extent to which the model (2.3) characterizes the data is

$$X^2 = SS(\sim F \sim) = X \sim F \sim = \sim \sim F'V_F^{-1} F \sim - \sim \sim b'(\sim X'V_F^{-1} X) b \sim \sim$$  \hspace{1cm} (2.6)$$

which has approximately a chi-square distribution with $D.F. = \{(\text{No. of rows in } X) - (\text{No. of columns in } X)\}$ in large samples, under the hypothesis that the model fits. For these data, $X^2 = 2.33$ with $D.F. = 6$ which is non-significant ($\alpha = .25$), and thus further consideration of this model is justified.

Finally, it can be argued that the test statistic (2.6) provides a means of partially checking the validity of the fundamental assumption that the incomplete data arose in a completely random manner. Caution, however, should be exercised in such interpretations since the 6 degrees of freedom in (2.6) here apply only to comparisons of corresponding marginal distributions in the three samples.

Since the model (2.3) adequately describes the data, tests of hypotheses with respect to the parameters comprising $\sim$ can be undertaken. In particular, for a general hypothesis of the form $H_0 : C \sim \sim = 0$ where $C$ is a known $(d \times 15)$ matrix of full rank, a suitable test statistic is

$$X^2 = SS(C \sim \sim = 0) = b' \sim C' [\sim C(X'V_F^{-1} X)^{-1} C']^{-1} \sim C b \sim \sim \hspace{1cm} (2.7)$$

which has approximately a chi-square distribution with $D.F. = d$ in large samples under $H_0$. An hypothesis which is of particular interest for these data is marginal homogeneity (marginal symmetry). This hypothesis may be written as

$$H_0 : \pi_j = \pi .j \text{ for } j = 1, 2, 3, 4 \hspace{1cm} (2.8)$$

which in terms of $\sim$ corresponds to

$$C \sim \sim = \begin{bmatrix}
1 & 1 & 1 & 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\
0 & -1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & -1 & 0 & 0 & 0 & -1 & 0 \\
0 & 0 & -1 & 0 & 0 & 0 & -1 & 0 & 1 & 1 & 0 & 0 & 0 & -1 & 0
\end{bmatrix} \sim \sim \hspace{1cm} (2.9)$$
For \( \zeta \) as specified in (2.9), the statistic (2.7) is \( \chi^2 = 11.41 \) with D.F. = 3, which is statistically significant (\( \alpha = 0.01 \)). When this hypothesis was investigated by GSK for the original data, ignoring the artificial data for the 1100 incompletely classified women, \( \chi^2 = 11.98 \) was obtained. In some sense this result is surprising since one would usually expect \( \chi^2 \)-statistics of this type to increase as more data were used in the analysis. However, here the additional data caused the differences between the respective estimates of the marginal probabilities for right eye and left eye to become smaller (see Table 2) than they were originally. On the other hand, Table 2 shows that the estimated

**TABLE 2**

ESTIMATED PROBABILITIES & STANDARD ERRORS FOR EYE DATA

<table>
<thead>
<tr>
<th>Cell</th>
<th>Data for Completely Classified Women (n=7477)</th>
<th>Data For All Women (n=8577)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( p )</td>
<td>( p )</td>
</tr>
<tr>
<td></td>
<td>s.e. (( p ))</td>
<td>s.e. (( \theta ))</td>
</tr>
<tr>
<td>(1,1)</td>
<td>.20329</td>
<td>.00465</td>
</tr>
<tr>
<td>(1,2)</td>
<td>.03557</td>
<td>.00214</td>
</tr>
<tr>
<td>(1,3)</td>
<td>.01658</td>
<td>.00148</td>
</tr>
<tr>
<td>(1,4)</td>
<td>.00889</td>
<td>.00108</td>
</tr>
<tr>
<td>(2,1)</td>
<td>.03129</td>
<td>.00219</td>
</tr>
<tr>
<td>(2,2)</td>
<td>.20222</td>
<td>.00464</td>
</tr>
<tr>
<td>(2,3)</td>
<td>.05777</td>
<td>.00270</td>
</tr>
<tr>
<td>(3,1)</td>
<td>.01043</td>
<td>.00118</td>
</tr>
<tr>
<td>(3,2)</td>
<td>.01564</td>
<td>.00144</td>
</tr>
<tr>
<td>(3,3)</td>
<td>.00481</td>
<td>.00248</td>
</tr>
<tr>
<td>(3,4)</td>
<td>.23699</td>
<td>.00492</td>
</tr>
<tr>
<td>(4,1)</td>
<td>.02741</td>
<td>.00189</td>
</tr>
<tr>
<td>(4,2)</td>
<td>.00481</td>
<td>.00080</td>
</tr>
<tr>
<td>(4,3)</td>
<td>.01096</td>
<td>.00120</td>
</tr>
<tr>
<td>(4,4)</td>
<td>.06580</td>
<td>.00287</td>
</tr>
<tr>
<td>(1,1)</td>
<td>.26428</td>
<td>.00510</td>
</tr>
<tr>
<td>(2,1)</td>
<td>.30173</td>
<td>.00530</td>
</tr>
<tr>
<td>(3,1)</td>
<td>.32847</td>
<td>.00543</td>
</tr>
<tr>
<td>(4,1)</td>
<td>.10552</td>
<td>.00355</td>
</tr>
<tr>
<td>(1,2)</td>
<td>.25505</td>
<td>.00504</td>
</tr>
<tr>
<td>(1,3)</td>
<td>.29738</td>
<td>.00529</td>
</tr>
<tr>
<td>(1,4)</td>
<td>.33529</td>
<td>.00546</td>
</tr>
<tr>
<td>(2,2)</td>
<td>.11248</td>
<td>.00365</td>
</tr>
</tbody>
</table>
standard errors for the estimates $\hat{\pi}$ of the cell probabilities derived using all the data are uniformly smaller than the estimated standard errors for $p$ which reflect only data for completely classified women; a similar statement applies to estimates of right eye and left eye marginal probabilities derived from $\hat{\pi}$ and $p$ respectively. Thus, by including the incompletely classified women in the analysis, more precise estimators of these parameters have been obtained in the sense of estimated standard error.

The above analysis has emphasized the use of the minimum-$\chi^2_1$ estimates $\hat{\pi}$. Alternatively, Reinfurt [1970] has implied that if the analysis here is repeated but with $V(\pi)$ estimated by $V(\hat{\pi})$ rather than $V(\hat{\pi}_G)$ and the process is continued in an iterative manner until successive estimates of $\pi$ are satisfactorily similar, the maximum likelihood estimate is obtained. However, this result as well as that of Hocking and Oxspring [1971] only applies to the unrestricted model (i.e., exclusive of any hypotheses on the underlying parameters). Maximum likelihood estimation of $\pi$ under hypotheses like (2.8) poses more difficult mathematical problems for situations where there is additional and incomplete categorical data of this type.

2.2 An incomplete split-plot experiment

Let us next consider a hypothetical experiment which has been undertaken to compare three drugs A, B, and C. Suppose the response of a given subject to any of these drugs can be observed in a quantal sense as either favorable or unfavorable. For $n_{ABC} = 46$ subjects, each of the three drugs was administered and the separate responses to each were noted; for $n_{AB} = 28$ subjects, only A and B were administered; for $n_A = 16$ subjects, only A. Groups of $n_{AC} = 25$, $n_{BC} = 26$, $n_B = 15$, $n_C = 14$ subjects were treated analogously. The observed results are given in Table 3.
TABLE 3
RESPONSES TO DRUGS A, B, C
(1 DENOTES FAVORABLE RESPONSE, 0 DENOTES UNFAVORABLE RESPONSE, AND * DENOTES THAT THE DRUG WAS NOT RECEIVED)

<table>
<thead>
<tr>
<th>PATTERNS OF RESPONSE</th>
<th>NUMBER OF RESPONDENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TOTAL = 170</td>
</tr>
<tr>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>*</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td></td>
</tr>
<tr>
<td>*</td>
<td></td>
</tr>
<tr>
<td>*</td>
<td></td>
</tr>
<tr>
<td>*</td>
<td></td>
</tr>
<tr>
<td>*</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td></td>
</tr>
<tr>
<td>*</td>
<td></td>
</tr>
<tr>
<td>*</td>
<td></td>
</tr>
<tr>
<td>*</td>
<td></td>
</tr>
<tr>
<td>*</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

For the 46 women who received all three drugs, the frequencies of the respective responses come from an illustrative example that has been treated by several authors including Cochran [1950], Bhapkar [1965], and GSK. The data for the other 124 subjects are artificial.
In this example, the incomplete nature of the data is considered to have been built into the experimental design. Thus, if the set of drugs a subject receives is determined randomly, it is appropriate to assume that the same parameters $\pi_A$, $\pi_B$, $\pi_C$ represent the probabilities of a favorable response to drugs $A$, $B$, $C$ respectively for each of the groups of subjects.

The data in this example can be analyzed in a manner similar to that demonstrated in Section 2.1. First of all, we view the results in Table 3 as coming from seven populations and arrange it in the array (2.10). Let

$$
\begin{bmatrix}
  n_{ABC}^i \\
  n_{AB}^i \\
  n_{AC}^i \\
  n_{BC}^i \\
  n_A^i \\
  n_B^i \\
  n_C^i \\
\end{bmatrix} =
\begin{bmatrix}
  6 & 16 & 2 & 4 & 2 & 4 & 6 & 6 \\
  12 & 4 & 4 & 8 \\
  5 & 10 & 4 & 6 \\
  4 & 12 & 5 & 5 \\
  10 & 6 \\
  11 & 4 \\
  5 & 9 \\
\end{bmatrix}
$$

(2.10)

$p_i = n_i/n_{..}$ where $i$ takes on the values $i = ABC, AB, AC, BC, A, B, C$, and let $\bar{z}_G$ be the compound vector $\bar{z}_G = (p_{ABC}^i, p_{AB}^i, p_{AC}^i, p_{BC}^i, p_A^i, p_B^i, p_C^i)$. If we then form the vector of linear functions $\bar{z} = A\bar{z}_G$ where $A$ is defined in (2.11),

$$
A =
\begin{bmatrix}
  1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
  1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
  1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
  0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
  0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
  0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
  0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
  0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
  0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
  0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
\end{bmatrix}
$$

(2.11)

then it follows that $\bar{z} = A\bar{z}_G$ where $\bar{z}$ is given in (2.12) and $\pi^i = (\pi_A^i, \pi_B^i, \pi_C^i)$. 
\[ E(F) = X\pi = \begin{bmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1 \\
1 & 0 & 0 \\
0 & 1 & 0 \\
1 & 0 & 0 \\
0 & 0 & 1 \\
0 & 1 & 0 \\
0 & 0 & 1 \\
\end{bmatrix} \begin{bmatrix}
\pi_A \\
\pi_B \\
\pi_C \\
\end{bmatrix} \] (2.12)

A consistent estimate \( \tilde{V}_F \) for the covariance matrix of \( F \) is given by \( \tilde{V}_F = A \tilde{V}(\tilde{p}_G) A' \) with \( \tilde{V}(\tilde{p}_G) \) being the block diagonal matrix shown in (2.13) which has

\[
\tilde{V}(\tilde{p}_G) = \begin{bmatrix}
V_{ABC}(p_{ABC}) & 0_{8,4} & 0_{8,4} & 0_{8,4} & 0_{8,2} & 0_{8,2} & 0_{8,2} \\
0_{4,2} & 0_{4,2} & 0_{4,2} & 0_{4,2} & 0_{4,2} & 0_{4,2} & 0_{4,2} \\
0_{4,2} & 0_{4,2} & 0_{4,2} & 0_{4,2} & 0_{4,2} & 0_{4,2} & 0_{4,2} \\
0_{2,2} & 0_{2,2} & 0_{2,2} & 0_{2,2} & 0_{2,2} & 0_{2,2} & 0_{2,2} \\
\end{bmatrix}
\] (2.13)

sub-matrices \( \tilde{V}_i(p_i) = (D \tilde{p}_i - p_i p_i')/n_i \) for \( i = ABC, AB, AC, BC, A, B, C \) respectively on the main diagonal. Following the approach discussed in the preceding section, the model (2.12) is fitted to the vector \( F \) by weighted least squares. The resulting estimates for \( \tilde{\pi}_A, \tilde{\pi}_B, \tilde{\pi}_C \) and their estimated standard errors as determined from
expressions analogous to (2.4) and (2.5) are given in the last column of Table 4.

**TABLE 4**

ESTIMATED PROBABILITIES AND (STANDARD ERRORS)

<table>
<thead>
<tr>
<th>PARAMETERS</th>
<th>SUBJECTS RECEIVING ALL THREE DRUGS (n=46)</th>
<th>SUBJECTS RECEIVING ONLY ONE OR TWO DRUGS (n=124)</th>
<th>ALL SUBJECTS (n=170)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \pi_A )</td>
<td>.609 (0.072)</td>
<td>.604 (0.058)</td>
<td>.607 (0.044)</td>
</tr>
<tr>
<td>( \pi_B )</td>
<td>.609 (0.072)</td>
<td>.629 (0.056)</td>
<td>.621 (0.044)</td>
</tr>
<tr>
<td>( \pi_C )</td>
<td>.348 (0.070)</td>
<td>.352 (0.059)</td>
<td>.350 (0.045)</td>
</tr>
<tr>
<td>( X^2 )-statistic (D.F. = 2)</td>
<td>6.58</td>
<td>12.03</td>
<td>18.79</td>
</tr>
<tr>
<td>( H_0: \pi_A = \pi_B = \pi_C )</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A test of the hypothesis \( H_0: \pi_A = \pi_B = \pi_C \) may be performed by using

\[
\mathbf{c} = \begin{bmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \end{bmatrix}
\]

in an expression analogous to (2.7). The resulting \( X^2 = 18.79 \) with D.F. = 2 is statistically significant (\( \alpha = 0.01 \)). Finally, it is worthwhile to note that the goodness of fit statistic analogous to (2.6) for the model (2.12) is \( X^2 = 1.37 \) with D.F. = 9; thus the assumption that the same parameters \( \pi_A, \pi_B, \pi_C \) represent the probabilities of a favorable response to drugs A, B, C respectively for each of the groups of subjects where they apply is consistent with the data.

For comparative purposes, results are also separately shown in Table 4 for the subjects receiving all three drugs and for the subjects receiving only one or two drugs. These were obtained by applying the previous type of analysis to the corresponding rows of (2.12); i.e., the first three rows for the former
and the last nine rows for the latter. For the hypothesis $H_0: \pi_A = \pi_B = \pi_C$
$X^2 = 6.58$ with D.F. = 2 and $X^2 = 12.03$ with D.F. = 2 respectively; and for goodness of fit, there is no test for the former and $X^2 = 1.29$ with D.F. = 6 for the latter.

2.3 An application with supplemental samples

This example is based on the data considered by Dyke and Patterson [1952], Bishop [1969], Goodman [1970], and other authors. Suppose in order to study the relationship between an individual's knowledge of cancer and extent of contact with reading materials, three samples were selected without replacement from a large population by simple random sampling (more complex sampling designs could be treated by the approach described in Johnson and Koch [1970] and Koch and Reinfurt [1970]). In sample I the individuals were cross classified according to the following three categorical variables:

i. whether they read newspapers or not,

ii. whether they read books or magazines (solid reading) or not,

iii. whether their knowledge of cancer was good or poor.

In sample II, subjects were classified only by (i) and (iii); in Sample III, only by (ii) and (iii).

The observed data are given in Table 5. For the 1729 subjects in sample I, the frequencies are the appropriate marginal totals of the data considered by Dyke and Patterson [1952] and other authors. The data for the other 910 subjects are artificial.

The incomplete nature of the data here is considered to have resulted from the structure of the sample survey design. Samples II and III might be separately conducted preliminary investigations involving variables (i) and (ii).

Alternatively, in many surveys it can be expected that response errors due to respondent
hostility or other causes will increase with the size of a questionnaire. In such situations it may be advantageous to split the questionnaire, resulting in data such as that of Table 5. In either case, the analytical strategy to be applied is similar to that of Section 2.2. However, for these data, attention will be directed at the relationship among the three categorical variables as it pertains to the conditional probability of good knowledge of cancer given fixed categories for exposure to newspapers and solid reading.

Following the approach described in Sections 2.1 and 2.2, we first view the results in Table 5 as coming from three populations and arrange it in an array as follows:

### TABLE 5

RESPONSES TO CANCER SURVEYS  
(* DENOTES UNMEASURED RESPONSE)

<table>
<thead>
<tr>
<th>PATTERNS OF RESPONSE</th>
<th>NUMBER OF RESPONDENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>NEWSPAPERS</strong></td>
<td><strong>SOLID READING</strong></td>
</tr>
<tr>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>YES</td>
<td>NO</td>
</tr>
<tr>
<td>YES</td>
<td>NO</td>
</tr>
<tr>
<td>NO</td>
<td>YES</td>
</tr>
<tr>
<td>NO</td>
<td>YES</td>
</tr>
<tr>
<td>NO</td>
<td>NO</td>
</tr>
<tr>
<td>NO</td>
<td>NO</td>
</tr>
<tr>
<td>YES</td>
<td>*</td>
</tr>
<tr>
<td>YES</td>
<td>*</td>
</tr>
<tr>
<td>NO</td>
<td>*</td>
</tr>
<tr>
<td>NO</td>
<td>*</td>
</tr>
<tr>
<td>*</td>
<td>YES</td>
</tr>
<tr>
<td>*</td>
<td>YES</td>
</tr>
<tr>
<td>*</td>
<td>NO</td>
</tr>
<tr>
<td>*</td>
<td>NO</td>
</tr>
</tbody>
</table>

\[n_1 = 1729\]

\[n_{II} = 340\]

\[n_{III} = 570\]
\[
\begin{bmatrix}
n_1^1 \\
n_1^2 \\
n_1^3
\end{bmatrix}
= \begin{bmatrix}
353 & 270 & 125 & 225 & 87 & 110 & 103 & 456 \\
90 & 100 & 40 & 110 \\
150 & 120 & 80 & 220
\end{bmatrix}
\]

Let \( p_1 = (n_1^1/n_1^1) \), \( p_1^2 = (n_1^2/n_1^2) \), and \( p_1^3 = (n_1^3/n_1^3) \) where \( n_1 = 1729 \), \( n_1^2 = 340 \), and \( n_1^3 = 570 \) are the sizes of the three samples. Let \( p_G \) be the augmented vector \( p_G' = (p_1', p_1^2', p_1^3') \). If we then form the vector of linear functions

\[
F = Ap_G
\]

where

\[
A = \begin{bmatrix}
1 & 7 & 0 & 7,1 & 0 & 7,3 & 0 & 7,1 & 0 & 7,3 & 0 & 7,1 \\
0 & 3,7 & 0 & 3,1 & 1 & 3 & 0 & 3,1 & 0 & 3,3 & 0 & 3,1 \\
0 & 3,7 & 0 & 3,1 & 0 & 3,3 & 0 & 3,1 & 1 & 3 & 0 & 3,1
\end{bmatrix}
\]

then it follows that \( E_F = X_{\pi_T} \), where \( X \) and \( \pi_T \) are given in (2.14). The

\[
E_F = X_{\pi_T} = \begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\
1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 1
\end{bmatrix}
\]

elements \( \pi_{nh} \) which make \( \pi_T \) may be interpreted as the probabilities that an individual is classified into the \( h \)-th category with respect to newspapers (1 if yes and 2 if no), the \( h \)-th category with respect to solid reading (1 if yes and 2 if no), and the \( j \)-th category with respect to knowledge of cancer (1 if good and
2 if poor). The estimated covariance matrix \( V_F \) for \( F \) is given by \( V_F = A[V(p_i^{(2)})]A' \), \( V(p_i^{(2)}) \) being the following block diagonal matrix

\[
V(p_i^{(2)}) = \begin{bmatrix}
V_I(p_i) & 0_{7,3} & 0_{7,3} \\
0_{3,3} & V_{II}(p_{II}) & 0_{3,3} \\
0_{3,3} & 0_{3,3} & V_{III}(p_{III})
\end{bmatrix}
\]

where \( V_I(p_i) = (D_{pi} - p_ip_i')/n_i \) for \( i = I, II, III \).

The vector \( \pi_T \) is estimated by applying weighted least squares to the model (2.14). The goodness of fit statistic analogous to (2.6) for the model (2.14) is \( \chi^2 = 1.02 \) with D.F. = 6. Hence, the model is consistent with the data. The resulting estimates \( \pi_T \) and their standard errors are shown in the last two columns of Table 6. In addition, \( \pi_{222} = (1 - j_7'\pi_T) \) and its standard error, where \( j_7 \) is a vector of ones, are included here. Finally, corresponding estimates from the Sample I data only are given for purposes of comparison.

### Table 6

**Estimated Probabilities for Dyke-Patterson Data**

<table>
<thead>
<tr>
<th>Pattern of Response</th>
<th>Solid</th>
<th>Cancer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Newspapers</td>
<td>Reading</td>
<td>Knowledge</td>
</tr>
<tr>
<td>Y</td>
<td>Y</td>
<td>G</td>
</tr>
<tr>
<td>Y</td>
<td>Y</td>
<td>F</td>
</tr>
<tr>
<td>Y</td>
<td>N</td>
<td>G</td>
</tr>
<tr>
<td>Y</td>
<td>N</td>
<td>F</td>
</tr>
<tr>
<td>N</td>
<td>Y</td>
<td>G</td>
</tr>
<tr>
<td>N</td>
<td>Y</td>
<td>P</td>
</tr>
<tr>
<td>N</td>
<td>N</td>
<td>G</td>
</tr>
<tr>
<td>N</td>
<td>N</td>
<td>P</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sample I</th>
<th>All Samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \hat{\pi} )</td>
<td>( \hat{\pi} )</td>
</tr>
<tr>
<td>s.e.( \hat{\pi} )</td>
<td>s.e.( \hat{\pi} )</td>
</tr>
<tr>
<td>.20416</td>
<td>.20433</td>
</tr>
<tr>
<td>.15616</td>
<td>.15529</td>
</tr>
<tr>
<td>.07230</td>
<td>.07278</td>
</tr>
<tr>
<td>.13013</td>
<td>.13024</td>
</tr>
<tr>
<td>.05032</td>
<td>.05127</td>
</tr>
<tr>
<td>.06362</td>
<td>.06284</td>
</tr>
<tr>
<td>.05957</td>
<td>.06105</td>
</tr>
<tr>
<td>.26374</td>
<td>.26220</td>
</tr>
</tbody>
</table>
In order to estimate the conditional probabilities of good knowledge of cancer for fixed categories with respect to newspapers and solid reading, several additional calculations are required. If we let \( \hat{\pi} \) be defined by \( \hat{\pi} = (\hat{\pi}_T, \hat{\pi}_{222}) \), then a consistent estimate for the covariance matrix for \( \hat{\pi} \) is

\[
\hat{\Sigma} = \left[ \begin{array}{c} I_7 \\ -J^{-1}_7 \\
\end{array} \right] \left[ X' \Sigma_{L} X \right]^{-1} \left[ \begin{array}{c} I_7 \\ -J^{-1}_7 \\
\end{array} \right].
\]

(2.15)

Let \( f_{hh'} = \hat{\pi}_{hh'} / (\hat{\pi}_{hh'} + \hat{\pi}_{hh''}) \) denote the estimated conditional probability for good knowledge of cancer for individuals in the \( h \)-th category with respect to newspapers and the \( h' \)-th category with respect to solid reading. Let \( \tilde{f} \) be the vector defined by \( \tilde{f}' = (f_{11}, f_{12}, f_{21}, f_{22}) \). The vector \( \tilde{f} \) may be written in matrix notation as \( \tilde{f} = \text{exp}(K \log_e (A_{\tilde{f}})) \) where \( A_{\tilde{f}} \) and \( K \) are displayed in (2.16);

\[
A_{\tilde{f}} \sim = \begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 \\
\end{bmatrix}
\]

and \( K \sim = \begin{bmatrix}
1 & -1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & -1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & -1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & -1 \\
\end{bmatrix}
\)

(2.16)

also, \( \log_e \) transforms a vector to the corresponding vector of logarithms and \( \text{exp} \) transforms a vector to the corresponding vector of anti-logarithms (i.e., exponential functions). A consistent estimate for the covariance matrix of \( \tilde{f} \) is

\[
\tilde{V}_f = D_{\tilde{f}} K D_{\tilde{f}} \Sigma_{\tilde{f}} V_{\tilde{f}} D_{\tilde{f}}^{-1} \Sigma_{\tilde{f}} V_{\tilde{f}} D_{\tilde{f}}^{-1}
\]

where \( D_{\tilde{f}} \) and \( D_{\tilde{g}} \) are diagonal matrices formed from the vectors \( \tilde{f} \) and \( \tilde{g} = A_{\tilde{f}} \tilde{f} \) respectively and \( \tilde{V} \) is the matrix in (2.15).
On applying these results to the data in this example, we obtain

\[
\hat{f} = \begin{bmatrix} .568 \\ .358 \\ .449 \\ .189 \end{bmatrix}
\quad \text{and} \quad \hat{V}_f = \begin{bmatrix} 2.937 & -2.289 & -1.651 & .039 \\ 5.485 & .133 & -2.424 \\ 11.120 & -.348 \\ 2.172 \end{bmatrix} \times 10^{-4} 
\]

In order to determine the influence that newspapers and solid readings have on the probability of good knowledge of cancer, we fit the model

\[
E(\hat{f}) = X_\hat{f} \hat{\beta} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & -1 & 1 \\ 1 & -1 & -1 \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix}
\]

(2.17)

by weighted least squares where \(\hat{\beta}_1\) is an overall mean, \(\hat{\beta}_2\) is a newspaper effect, and \(\hat{\beta}_3\) is a solid reading effect. The resulting estimate \(\hat{\beta}\) of \(\beta\) and its estimated covariance matrix \(\hat{V}_\beta\) are

\[
\hat{\beta} = \begin{bmatrix} .385 \\ .076 \\ .115 \end{bmatrix}
\quad \text{and} \quad \hat{V}_\beta = \begin{bmatrix} .830 & -.101 & .228 \\ -1.275 & -.623 \\ 1.220 \end{bmatrix} \times 10^{-4}
\]

The goodness of fit test analogous to (2.6) for the model (2.15) is \(X^2 = 1.01\) with D.F. = 1. Thus, the model is supported by the data, which implies that there is no interaction between the effects of newspapers and solid reading with respect to the vector \(\hat{f}\). Finally, for the hypothesis \(H_0 : \hat{\beta}_2 = 0\), \(X^2 = 44.74\) with D.F. = 1; and for the hypothesis \(H_0 : \hat{\beta}_3 = 0\), \(X^2 = 108.50\) with D.F. = 1. Thus, both of these main effects are statistically significant (\(\alpha = .01\)).
If only the data in sample I are considered, the results corresponding
to $b$ and $\tilde{V}_b$ are

$$
\begin{bmatrix}
0.382 \\
0.078 \\
0.115
\end{bmatrix}
\quad \text{and} \quad
\begin{bmatrix}
1.230 & -0.076 & 0.308 \\
1.480 & -0.695 & 1.553
\end{bmatrix} \times 10^4
$$

Similarly, the goodness of fit test statistic as well as those for $H_0: \beta_2 = 0$
and $H_0: \beta_3 = 0$ are $X^2 = 0.89$, $X^2 = 40.92$ and $X^2 = 84.45$, respectively, each
with D.F. = 1. Thus, it is apparent that inclusion of Samples II and III in
the analysis produced some improvement in the precision of the estimate of $\bar{b}$.

2.4 A Graeco-Latin square split-plot experiment

The data in Table 7 represent hypothetical results of a typical experi-
ment conducted on three specially instrumented cars by the University of North
Carolina Highway Safety Research Center. The objective here is to determine the
effect of three different driver training methods and three different examining
procedures on an individual's ability to pass a driving skill test. Representa-
tive subjects from populations corresponding to the three driver training methods
participate in a driving skill test at three different times under three different
combinations of car and examining procedure. These combinations are determined
according to the Graeco-Latin square in Table 8, where rows represent the driver
training method; columns, the three trials; Latin letters, the three cars; and
Greek letters, the three testing procedures. In some cases, certain subjects do
TABLE 7

RESULTS OF DRIVING TESTS
(P = PASS, F = FALL, * = NOT TAKEN)

<table>
<thead>
<tr>
<th>PATTERNS OF RESPONSE</th>
<th>NUMBER OF RESPONDENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRIAL</td>
<td>GROUP</td>
</tr>
<tr>
<td></td>
<td>A</td>
</tr>
<tr>
<td>I</td>
<td>II</td>
</tr>
<tr>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>P</td>
<td>F</td>
</tr>
<tr>
<td>F</td>
<td>P</td>
</tr>
<tr>
<td>F</td>
<td>P</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>P</td>
<td>*</td>
</tr>
<tr>
<td>P</td>
<td>*</td>
</tr>
<tr>
<td>F</td>
<td>*</td>
</tr>
<tr>
<td>F</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>P</td>
<td>F</td>
</tr>
<tr>
<td>F</td>
<td>P</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

TABLE 8

DESIGN OF INSTRUMENTED CAR EXPERIMENT

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Xα</td>
<td>Yβ</td>
</tr>
<tr>
<td>B</td>
<td>Zβ</td>
<td>Xγ</td>
</tr>
<tr>
<td>C</td>
<td>Yγ</td>
<td>Zα</td>
</tr>
</tbody>
</table>

\[ Zγ \]
not participate in all three tests either because of time scheduling problems or because of cost considerations. Thus, the data in Table 7 includes groups of subjects with incomplete response vectors which may be viewed as having arisen either in the context of the missing data in Section 2.1 or the supplemental sample data in Section 2.3.

As in the other examples, we first view the results in Table 7 as coming from nine populations and thus formulate the array displayed in (2.18).

\[
\begin{bmatrix}
\begin{array}{ccccc}
\mathbf{p}_{A,1} & 15 & 24 & 9 & 15 & 5 & 13 & 10 & 9 \\
\mathbf{p}_{B,1} & 12 & 3 & 22 & 9 & 16 & 13 & 13 & 12 \\
\mathbf{p}_{C,1} & 6 & 14 & 6 & 6 & 10 & 25 & 13 & 20 \\
\mathbf{p}_{A,2} & 6 & 21 & 8 & 5 & & & & \\
\mathbf{p}_{B,2} & 14 & 8 & 10 & 8 & & & & \\
\mathbf{p}_{C,2} & 3 & 6 & 11 & 20 & & & & \\
\mathbf{p}_{A,3} & 18 & 10 & 7 & 5 & & & & \\
\mathbf{p}_{B,3} & 8 & 11 & 8 & 13 & & & & \\
\mathbf{p}_{C,3} & 5 & 5 & 18 & 12 & & & & \\
\end{array}
\end{bmatrix}
\] (2.18)

For each \( n_i \) in (2.18) define \( \mathbf{p}_i = (n_i/n) \) where \( i \) takes on values A1, B1, ..., C3 and \( n_{A,1} = n_{B,1} = n_{C,1} = 100, n_{A,2} = n_{B,2} = n_{C,2} = n_{A,3} = n_{B,3} = n_{C,3} = 40 \) are the sums of the elements in \( \mathbf{p}_{A,1}, \mathbf{p}_{B,1}, \mathbf{p}_{C,1}, \mathbf{p}_{A,2}, \mathbf{p}_{B,2}, \mathbf{p}_{C,2}, \mathbf{p}_{A,3}, \mathbf{p}_{B,3}, \mathbf{p}_{C,3} \) respectively. Let \( \mathbf{p}_G \) be the compound vector \( \mathbf{p}_G = (\mathbf{p}_{A,1}, \mathbf{p}_{B,1}, \mathbf{p}_{C,1}, \mathbf{p}_{A,2}, \mathbf{p}_{B,2}, \mathbf{p}_{C,2}, \mathbf{p}_{A,3}, \mathbf{p}_{B,3}, \mathbf{p}_{C,3}) \). If we form the vector of linear functions \( \mathbf{f} = \mathbf{a}_G \) where \( \mathbf{a} \) is defined in (2.19), then the resulting elements of \( \mathbf{f} \) represent estimated probabilities with which subjects in the various groups pass the driving test.
on respective trials. Under the assumption that there is no second or higher order interaction among the four factors training method, trial, car, and examining procedure with respect to these probabilities of passing, then it follows that $F(x) = x^\beta$ where $x$ and $\beta$ are displayed in (2.20). The component
elements of $\beta$ are an overall mean $\mu$, training method effects $\tau_1$ and $\tau_2$, trial effects $\rho_1$ and $\rho_2$, car effects $\gamma_1$ and $\gamma_2$, and examination effects $\xi_1$ and $\xi_2$. A consistent estimate for the covariance matrix of $F$ is $\tilde{V}_F = A[V(P_G)]A'$ where $V(P_G)$ is a block diagonal matrix with the matrices $V(P_i) = [D_i - p_i P_i]/n_i$ for $i = A1, B1, C1, A2, B2, C2, A3, B3, C3$ respectively being the main diagonal blocks.

As with the previous examples, the model (2.20) can be fitted to the
vector $\mathbf{F}$ by weighted least squares. The resulting estimates $\mathbf{b}$ of $\beta$ and their standard errors are given in the last two columns of Table 9. Similarly,

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Subjects in All Three Trials (n=300)</th>
<th>Subjects in Two Trials Only (n=240)</th>
<th>All Subjects (n=540)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu$</td>
<td>0.482</td>
<td>0.481</td>
<td>0.482</td>
</tr>
<tr>
<td>$\tau_1$</td>
<td>0.076</td>
<td>0.094</td>
<td>0.088</td>
</tr>
<tr>
<td>$\tau_2$</td>
<td>-0.028</td>
<td>-0.022</td>
<td>-0.029</td>
</tr>
<tr>
<td>$\rho_1$</td>
<td>0.026</td>
<td>0.051</td>
<td>0.033</td>
</tr>
<tr>
<td>$\rho_2$</td>
<td>-0.038</td>
<td>-0.053</td>
<td>-0.040</td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>0.016</td>
<td>0.003</td>
<td>0.011</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>-0.024</td>
<td>-0.005</td>
<td>-0.015</td>
</tr>
<tr>
<td>$\xi_1$</td>
<td>0.099</td>
<td>0.153</td>
<td>-0.121</td>
</tr>
<tr>
<td>$\xi_2$</td>
<td>0.022</td>
<td>-0.014</td>
<td>-0.009</td>
</tr>
</tbody>
</table>

$X^2$-tests for the effects of the respective factors and the adequacy of the model are determined as in (2.6) and (2.7). These results are displayed in the last column of Table 10. For purposes of comparison, analogous statistics are also

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>$X^2$-statistic Subjects in All Three Trials</th>
<th>$X^2$-statistic Subjects in Two Trials Only</th>
<th>$X^2$-statistic All Subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\tau_1=\tau_2=0$</td>
<td>11.22</td>
<td>9.95</td>
<td>25.50</td>
</tr>
<tr>
<td>$\rho_1=\rho_2=0$</td>
<td>2.56</td>
<td>2.77</td>
<td>4.61</td>
</tr>
<tr>
<td>$\gamma_1=\gamma_2=0$</td>
<td>1.34</td>
<td>0.02</td>
<td>0.79</td>
</tr>
<tr>
<td>$\xi_1=\xi_2=0$</td>
<td>33.84</td>
<td>25.24</td>
<td>62.52</td>
</tr>
<tr>
<td>Residual</td>
<td>0.00</td>
<td>0.58</td>
<td>4.52</td>
</tr>
</tbody>
</table>
separately shown in Tables 9 and 10 for the subjects participating in all three trials and for the subjects participating in only two of the trials. These values were obtained by applying the methods described here to the appropriate rows of (2.20), i.e., the first nine rows for the former case and the last twelve rows for the latter.

All of the results in Table 10 indicate that car effects and trial effects are not important. Hence, a revised model can be fitted to $\bar{x}$ with these factors excluded (i.e., columns 4, 5, 6, 7 are deleted from $X$ and $\rho_1$, $\rho_2$, $\gamma_1$, $\gamma_2$ are deleted from $\beta$). The corresponding estimates and test statistics for this analysis of the data on all subjects are given in Table 11. Finally, it

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>s.e.</th>
<th>Hypothesis</th>
<th>$X^2$</th>
<th>D.F.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu$</td>
<td>.480</td>
<td>.013</td>
<td>Residual</td>
<td>9.80</td>
<td>16</td>
</tr>
<tr>
<td>$\tau_1$</td>
<td>.087</td>
<td>.018</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tau_2$</td>
<td>-.031</td>
<td>.019</td>
<td>$\tau_1=\tau_2=0$</td>
<td>24.81</td>
<td>2</td>
</tr>
<tr>
<td>$\xi_1$</td>
<td>.119</td>
<td>.019</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\xi_2$</td>
<td>.007</td>
<td>.018</td>
<td>$\xi_1=\xi_2=0$</td>
<td>61.72</td>
<td>2</td>
</tr>
</tbody>
</table>

is worthwhile noting that if it can be a priori assumed that there are no differences among cars and no differences among trials, then the parameters $\rho_1$, $\rho_2$, $\gamma_1$, $\gamma_2$ correspond to the (training method x examination procedure) interaction effects. In this context, these results imply that the main effects of training method and examination procedure are statistically significant ($\alpha=.01$), and that the corresponding second order interaction is negligible.
2.5 A growth curve problem

In many areas of research, longitudinal data are collected from subjects at several different points in time. The following example is of interest with respect to the analysis of categorical data arising from such investigations.

Suppose four different diets (I, II, III, IV) designed to reduce blood cholesterol are assigned to four different groups of people. At the end of each of three time periods, a blood sample is taken from each available person and classified as to whether the blood cholesterol is normal or abnormal. For diets II, III, and IV, the subjects were closely supervised (e.g., within a hospital or clinic environment), and responses for all three time periods were obtained. Subjects on diet I were not as closely supervised, however, and for many of these individuals blood sample data were obtainable only for one or two of the time periods. The resulting artificial data are shown in Table 12. It is assumed that the incomplete data in this situation occurred because of scheduling problems (e.g., subjects in group I lived at home and were not able to attend all three appointments for blood samples either because of other commitments, unrelated illness, travel distance, etc.) and thus may be viewed either in the context of the missing data in Sections 2.1 or 2.4 or the supplemental sample data in Section 2.3.

Let us now arrange the results in Table 12 in the array displayed in (2,21).

\[
\begin{bmatrix}
\mathbf{n}_{I,1} & 2 & 2 & 8 & 9 & 9 & 15 & 27 & 28 \\
\mathbf{n}_{II,1} & 7 & 2 & 5 & 2 & 31 & 5 & 32 & 6 \\
\mathbf{n}_{III,1} & 16 & 13 & 9 & 3 & 14 & 4 & 15 & 6 \\
\mathbf{n}_{IV,1} & 31 & 0 & 6 & 0 & 22 & 2 & 9 & 0 \\
\mathbf{n}_{I,2} & 1 & 4 & 6 & 14 \\
\mathbf{n}_{I,3} & 1 & 3 & 7 & 14 \\
\mathbf{n}_{I,4} & 3 & 4 & 8 & 10 \\
\mathbf{n}_{I,5} & 6 & 19 \\
\end{bmatrix}
\]
### TABLE 12

**RESULTS OF BLOOD CHOLESTEROL TESTS**  
(N = NORMAL, A = ABNORMAL, * = NOT TAKEN)

<table>
<thead>
<tr>
<th>PATTERNS OF RESPONSE</th>
<th>NUMBER OF RESPONDENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TIME PERIOD</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>N</td>
<td>A</td>
</tr>
<tr>
<td>N</td>
<td>A</td>
</tr>
<tr>
<td>A</td>
<td>N</td>
</tr>
<tr>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>A</td>
<td>A</td>
</tr>
</tbody>
</table>

\[ n_{II,1}=90 \quad n_{III,1}=80 \quad n_{IV,1}=70 \]

| N                    | N | * | 1  |
| N                    | A | * | 4  |
| A                    | A | * | 6  |

\[ n_{I,2}=25 \]

| N                    | * | N | 1  |
| N                    | * | A | 3  |
| A                    | * | N | 7  |

\[ n_{I,3}=25 \]

| *                    | N | N | 3  |
| *                    | N | A | 4  |
| *                    | A | N | 8  |

\[ n_{I,4}=25 \]

| *                    | A | A | 10 |

\[ n_{I,5}=25 \]
Let $p_i = (n_i / n)$ for $i = I, III, IV, II$, respectively where $n_i$ denotes the sum of the elements in the vector $n_i$; i.e., $n_{I,1} = 100$, $n_{II,1} = 90$, $n_{III,1} = 80$, $n_{IV,1} = 70$, $n_{I,2} = n_{I,3} = n_{I,4} = n_{I,5} = 25$. Let $p_G$ be the compound vector

$$p_G' = (p_{I,1}' P_{II,1}' P_{III,1}' P_{IV,1}' P_{I,2}' P_{I,3}' P_{I,4}' P_{I,5}')$$  (2.22)

If we form the vector of linear functions $\mathbf{F} = \mathbf{A} \mathbf{p}_G$ where $\mathbf{A}$ is defined in (2.23), then the resulting elements of $\mathbf{F}$ represent the estimated probabilities of normal and abnormal blood cholesterol for the respective time periods in the four groups of subjects. If the trend according to which the probability of normal
blood cholesterol changes over time is a logistic curve, then the logit function vector \( \tilde{f} \) obtained as \( \tilde{f} = \tilde{K} \log_{\tilde{e}} (\tilde{F}) \) becomes of interest, where \( \log_{\tilde{e}} (\tilde{F}) \) is the vector of logarithms of elements in \( \tilde{F} \), and \( \tilde{K} \) is a 19 \times 38 block diagonal matrix with nineteen submatrices of the type \( \tilde{K}_1 = [1 -1] \) forming the main diagonal. A consistent estimate for the covariance matrix of \( \tilde{F} \) is

\[
V_{\tilde{F}} = \tilde{K} D_{\tilde{F}}^{-1} A \{ V(p_i) \} A' D_{\tilde{F}}^{-1} \tilde{K} \quad (2.24)
\]

where \( V(p_i) \) is a block diagonal matrix with the matrices \( V(p_i) = \{ D_{\tilde{p}_i} - p_i p_i' \}/n_i \) for \( i = \text{II, III, IIII, IVI, II} \), \( \text{I2, I3, I4, I5} \) respectively representing the main diagonal blocks.

In order to compare the groups with respect to trends in their blood cholesterol, we fit the model given by (2.25) to the vector \( \tilde{f} \) by weighted least squares where \( \beta_{0,i} \) and \( \beta_{1,i} \) represent the intercept and slope parameters for the

\[
\begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 2 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 & 0 \\
0 & 0 & 1 & 2 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 1 & 0 \\
0 & 0 & 0 & 0 & 1 & 2 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 1 & 2 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 2 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 2 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}
\begin{bmatrix}
\beta_{0,I} \\
\beta_{1,I} \\
\beta_{0,II} \\
\beta_{1,II} \\
\beta_{0,III} \\
\beta_{1,III} \\
\beta_{0,IV} \\
\beta_{1,IV}
\end{bmatrix}
\]

(2.25)
growth curve fitted to the i-th group i = I,II,III,IV. The resulting estimate b of β and the estimated standard errors are shown in Table 13. The

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>s.e.</th>
</tr>
</thead>
<tbody>
<tr>
<td>β₀[I]</td>
<td>-1.400</td>
<td>.165</td>
</tr>
<tr>
<td>β₁[I]</td>
<td>.549</td>
<td>.124</td>
</tr>
<tr>
<td>β₀[II]</td>
<td>-1.548</td>
<td>.252</td>
</tr>
<tr>
<td>β₁[II]</td>
<td>1.570</td>
<td>.207</td>
</tr>
<tr>
<td>β₀[III]</td>
<td>.045</td>
<td>.221</td>
</tr>
<tr>
<td>β₁[III]</td>
<td>.335</td>
<td>.170</td>
</tr>
<tr>
<td>β₀[IV]</td>
<td>.043</td>
<td>.229</td>
</tr>
<tr>
<td>β₁[IV]</td>
<td>1.441</td>
<td>.249</td>
</tr>
</tbody>
</table>

The goodness of fit test for the model (2.25) yields $X^2 = 3.61$ with D.F. = 11. Thus, the model is consistent with the data. Results for tests of hypotheses corresponding to pairwise comparisons of the intercepts and slopes in the respective groups have been determined in a manner analogous to (2.7) and are given in Table 14. Hence, it is apparent that treatments I and III have similar time trends as do II and IV. Also treatments I and II have similar initial effects as

<table>
<thead>
<tr>
<th>DIETS COMPARED</th>
<th>INTERCEPTS</th>
<th>SLOPES</th>
</tr>
</thead>
<tbody>
<tr>
<td>I-II</td>
<td>0.24</td>
<td>17.81</td>
</tr>
<tr>
<td>I-III</td>
<td>27.39</td>
<td>1.03</td>
</tr>
<tr>
<td>I-IV</td>
<td>26.06</td>
<td>10.26</td>
</tr>
<tr>
<td>II-III</td>
<td>22.58</td>
<td>21.21</td>
</tr>
<tr>
<td>II-IV</td>
<td>21.81</td>
<td>0.16</td>
</tr>
<tr>
<td>III-IV</td>
<td>0.00</td>
<td>13.45</td>
</tr>
</tbody>
</table>
do III and IV. However, the other comparisons are statistically significant (α=.01). These conclusions may be summarized by fitting the revised model displayed in (2.26) where $\beta_1^*$ and $\beta_2^*$ are the two distinct intercept parameters and

$$E(\tilde{f}) = X_0^\beta^* = $$

\[
\begin{bmatrix}
1 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 \\
1 & 0 & 2 & 0 \\
1 & 0 & 0 & 1 \\
1 & 0 & 0 & 2 \\
0 & 1 & 0 & 0 \\
0 & 1 & 1 & 0 \\
0 & 1 & 2 & 0 \\
0 & 1 & 0 & 0 \\
0 & 1 & 0 & 1 \\
0 & 1 & 0 & 2 \\
1 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 \\
1 & 0 & 0 & 0 \\
1 & 0 & 2 & 0 \\
1 & 0 & 1 & 0 \\
1 & 0 & 2 & 0 \\
1 & 0 & 0 & 0
\end{bmatrix}^\beta^* 
\]

(2.26)

$\beta_3^*$ and $\beta_4^*$ are the two distinct slope parameters. On fitting the model (2.26) by weighted least squares, we obtain estimates $\hat{\beta}^*$ and estimated covariance-matrix $V_{\hat{\beta}^*}$ as shown in (2.27). The goodness of fit test for the model (2.26) yields $X^2 = 5.69$ with D.F. = 15 which implies that this model also is consistent with the data. Finally, for tests of the hypotheses $H_0 : \beta_1^* = \beta_2^*$ and $H_0 : \beta_3^* = \beta_4^*$, $X^2 = 89.74$ with D.F. = 1 and $X^2 = 66.09$ with D.F. = 1 respectively. The results are:
3. STATISTICAL THEORY

The methodology which has been applied in the preceding examples is very similar to that given by GSK for the complete data case. In particular, we assume that there are $s$ populations of elements from which independent random samples of fixed sizes $n_1, n_2, \ldots, n_s$, respectively are selected. The responses of the $n_i$ elements from the $i$-th population are classified into $r_i$ categories, with $n_{ij}$, where $j=1,2,\ldots,r_i$, denoting the number of elements classified into the $j$-th response category for the $i$-th population. This structure is different from that considered in GSK because it allows the definition of response categories as well as their number (i.e., the $r_i$) to vary from one population to another.

The vector $n_i$, where $n_i = (n_{i1}, n_{i2}, \ldots, n_{ir_i})$, will be assumed to follow the multinomial distribution with parameters $n_i$ and $\pi_i$, where $\pi_{ij}$ is the probability of response $j$ in population $i$. Thus, the relevant product multinomial model is

$$
\phi = \prod_{i=1}^{s} \left\{ \frac{n_{i1}! \cdot \cdots \cdot n_{ir_i}!}{r_i \prod_{j=1}^{r_i} n_{ij}!} \frac{r_i}{\pi_{ij}} \right\}
$$

(3.1)

Other models like those considered by Johnson and Koch [1970] are applicable for more complex sampling situations.
Let \( \mathbf{p}_1 = (n_1/n_1) \) and let \( \mathbf{p}_G \) be the compound vector defined by \( \mathbf{p}_G = (p_1^1, p_2^1, \ldots, p_s^1) \). A consistent estimate for the covariance matrix of \( \mathbf{p}_G \) is given by the block diagonal matrix \( \mathbf{y}(\mathbf{p}_G) \) with the matrices \( \mathbf{y}_1(\mathbf{p}_1) = \frac{1}{n_1} \{ \mathbf{D} \mathbf{p}_1 - \mathbf{p}_1 \mathbf{p}_1' \} \) for \( i=1,2,\ldots,s \) representing the main diagonal; here \( \mathbf{D} \) is a diagonal matrix with elements in the vector \( \mathbf{p}_1 \) on the main diagonal.

Let \( \mathbf{F}_1(\mathbf{p}_G), \mathbf{F}_2(\mathbf{p}_G), \ldots, \mathbf{F}_u(\mathbf{p}_G) \) be a set of \( u \) functions of \( \mathbf{p}_G \), each with partial derivatives up to order two with respect to the elements of \( \mathbf{p}_G \) existing throughout a region containing \( \tilde{\mathbf{p}}_G = \mathbf{E}(\mathbf{p}_G) \) where \( \tilde{\mathbf{p}}_G = (\tilde{\mathbf{p}}_1, \tilde{\mathbf{p}}_2, \ldots, \tilde{\mathbf{p}}_s) \).

If \( \tilde{\mathbf{y}} = \tilde{\mathbf{y}}(\mathbf{p}_G) \) is defined by \( \tilde{\mathbf{y}}'(\mathbf{p}_G) = (\mathbf{F}_1(\mathbf{p}_G), \mathbf{F}_2(\mathbf{p}_G), \ldots, \mathbf{F}_u(\mathbf{p}_G)) \), then the covariance matrix of \( \tilde{\mathbf{y}} \) can be consistently estimated by \( \tilde{\mathbf{y}}_{\mathbf{F}} = \mathbf{H}^*(\mathbf{y}(\mathbf{p}_G)) \mathbf{H}^* \) where

\[
\mathbf{H} = \left[ \frac{\partial \tilde{\mathbf{y}}(x)}{\partial x} \right]_{\tilde{\mathbf{y}} = \tilde{\mathbf{y}}(\mathbf{p}_G)}.
\]

In all applications, the functions comprising \( \tilde{\mathbf{y}} \) are chosen so that \( \tilde{\mathbf{y}}_{\mathbf{F}} \) is asymptotically non-singular.

The function vector \( \tilde{\mathbf{y}} \) is a consistent estimator of \( \tilde{\mathbf{y}}(\mathbf{p}_G) \). Hence, consideration can then be directed at fitting a linear model \( \tilde{\mathbf{y}}(\mathbf{p}_G) = \mathbf{X} \hat{\beta}, \) where \( \mathbf{X} \) is a known \((\hat{u} \times t)\) coefficient matrix and \( \hat{\beta} \) is an unknown \((t \times 1)\) parameter vector. Weighted least squares is applied to determine a BAN estimator \( \hat{\beta} \) for \( \hat{\beta} \) as indicated in (2.4). Statistical tests for the fit of the model and for linear hypotheses involving \( \hat{\beta} \) can be undertaken by applying (2.6) and (2.7).

For the examples considered in Section 2, primary emphasis was placed on cases where \( \tilde{\mathbf{y}}(\mathbf{p}_G) \) were linear functions obtained as \( \tilde{\mathbf{y}}(\mathbf{p}_G) = \mathbf{A} \mathbf{p}_G \); in this event \( \mathbf{y}_{\mathbf{F}} = \mathbf{A}^* \mathbf{y}(\mathbf{p}_G) \mathbf{A}^* \). The functions comprising \( \tilde{\mathbf{y}} \) in these examples were either estimates of cell probabilities or marginals thereof. Thus, even though the response categories for the respective s populations were not necessarily the same, they were always related in the sense of involving the same underlying parameters. This fact defines the underlying principle by means of which information on both completely classified subjects and incompletely
classified subjects can be synthesized together in the analysis. Finally, it should be indicated that in certain situations like that illustrated in Section 2.3, a two-stage approach is required. At the first stage, the methods which have been described here are used to determine estimates \( \hat{\pi} \) of \( \pi \) which reflect all the available data. Corresponding to \( \hat{\pi} \), an estimated covariance matrix \( \hat{\Sigma} \) is also determined. Then, at the second stage, \( \hat{\pi} \) and \( \hat{\Sigma} \) play the same role as \( \hat{p}_G \) and \( \hat{\Sigma}(\hat{p}_G) \) do at the first stage. In other words, if \( f(\pi) \) represents a set of functions which are of interest, then corresponding estimates are given by \( f(\hat{\pi}) \) with estimated covariance matrix \( \hat{\Sigma}_f = \hat{H} \hat{\Sigma} \hat{H}' \) where \( \hat{H} = \left[ df(x)/dx \big| x = \hat{\pi} \right] \).

Finally, suitable linear models \( E\{f(\hat{\pi}) \} = \sum_{\pi} f(\pi) = \sum_{\pi} \Sigma_{\pi} \) are then fitted by weighted least squares.

Essentially all the analyses described in this paper can be undertaken by using the same computer program cited in GSK. The only refinements required are suitable modifications of the relevant matrix operations. Also, in some cases, if certain \( n_{ij} \) are zero, it may be necessary to replace them by \( (1/n_{ij}) \) in order to prevent \( \Sigma_f \) from being singular (See GSK and Berkson [1955] for more details in this respect).

In summary, a straightforward generalization of the GSK approach has been formulated for situations involving incompletely classified categorical data. The corresponding methodology was shown to provide the analyst with a powerful tool for dealing with a wide variety of problems. Thus, in some sense, one can say that incomplete categorical data can be more readily handled than incomplete continuous data provided that the sample size is large (e.g., as in the examples reported here). In this context, Kleinbaum [1970] discusses how Wald statistics can be applied to incomplete continuous data arising in
situations like those emphasized here. Alternative methods have been discussed by Afifi and Elashoff [1966, 1967, 1969a, 1969b], Hocking and Smith [1968], and Hartley and Hocking [1971]. However, if computational difficulties arise in the application of any of these methods, then it may be more convenient to categorize the continuous data corresponding to such cases and apply the analysis reported here.

REFERENCES


Dyke, G.V. and Patterson, H.D. [1952]. Analysis of factorial arrangements when the data are proportions. *Biometrics* 8, 1-12.


John, P.W.M. [1966]. Augmenting 2^{n-1} designs. Technometrics 8, 469-80.


