This research was supported in part by National Institutes of Health, Institute of General Medical Sciences through Grant GM-70004-04 and by the U. S. Bureau of the Census through Joint Statistical Agreement JSA 74-2.

AN APPLICATION OF THE CLUMPED BINOMIAL MODEL TO THE ANALYSIS OF CLUSTERED ATTRIBUTE DATA

By

Gary G. Koch, H. Dennis Tolley, and Jean L. Freeman

Department of Biostatistics
University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series No. 990

JANUARY 1975

AN APPLICATION OF THE CLUMPED BINOMIAL MODEL TO THE ANALYSIS OF CLUSTERED ATTRIBUTE DATA

by

Gary G. Koch, H. Dennis Tolley, and Jean L. Freeman

Department of Biostatistics
School of Public Health
University of North Carolina
Chapel Hill, North Carolina, 27514 U.S.A.

SUMMARY

This paper is concerned with the statistical analysis of situations where the experimental units are clusters of subjects, each of whom is classified with respect to the presence of a particular (binary response) attribute. For this purpose, the clumped binomial probability model is used to define an underlying mean rate parameter in terms of which the effects of certain treatments as well as cluster size itself can be investigated. The resulting methodology involves a synthesis of certain maximum likelihood and weighted least squares procedures; it is illustrated for a three-dimensional contingency table pertaining to numbers of deaths in litters of baby mice.

1. INTRODUCTION

Recent research papers (Grizzle, et al. [1969] and Berkson [1968]) have used the data originally presented by Kastenbaum and Lamphiear [1959] to illustrate tests of no second order interaction with respect to a multiplicative model for a three-dimensional contingency table. In the experiment reported by Kastenbaum and Lamphiear (see Table 1), the number of deaths of

TABLE 1

DEPLETION DATA OF KASTENBAUM AND LAMPHIEAR [1959]

Litter Size	Treatment	Number of Depletions			
		0	1	2+	Total
7	A	58	11	5	74
	B	75	19	7	101
8	A	49	14	10	73
	B	58	17	8	83
9	A	33	18	15	66
	B	45	22	10	77
10	A	15	13	15	43
	B	39	22	18	79
11	A	4	12	17	33
	B	5	15	8	28

baby mice prior to weaning was observed for two different treatments, A and B, over several litter sizes. Each litter was recorded to have zero, one, or two or more depletions. The analyses which have been previously reported in the literature for these data were concerned with the effects of treatment and litter size on depletions. In particular, their results have suggested

that there is no apparent interaction between these factors.

This example actually belongs to a broad family of statistical problems. A variety of situations exist where the effects of treatments on experimental units which are clusters of subjects interact with cluster size with respect to the probability with which such subjects exhibit a particular (binary response) attribute. For example, the number of children in a family may be an important determinant of a child's physical and mental health. In this case, interaction between family size and socioeconomic status may be important. Another type of example would include studies in which small classes of students are taught a skill or receive therapeutic counsel by one of several techniques. Evaluation of a technique is by failure to learn, and the interaction of technique to class size may be of considerable importance. Similar examples such as the results of replicated diagnostic tests on individual patients in clinical trials and load size of case workers all illustrate a common statistical question.

In many studies like those previously mentioned, the occurrence of events like death or failure for each cluster are summarized in a contingency table like that shown in Table 1 for the mice depletion example. Here, the cases where there were two or more events are "clumped" together into a common category. One reason for clumping is to reduce the size of the table as well as to provide a format in which the response categories are the same for all the treatment x cluster size sub-populations. Secondly, if there are potential errors in accurately determining the actual number of events for a given cluster whenever it is in excess of a particular upper bound because of time and cost constraints, then clumping can be regarded as a data editing procedure which permits the effects of any resulting measurement bias to be bypassed while often still retaining much (and occasionally most) of the information from the data. Finally, some data are simply collected in a

clumped data format because of the nature of the measurement technique or because "many failures" is of no greater relevance to an experimenter than "at least r failures."

The purpose of this paper is to illustrate a procedure for the statistical analysis of treatment x cluster size interaction for (binary response) attribute data. Such interaction will be investigated in terms of a mean score type function which is determined separately for each treatment x cluster size sub-population. The relationships between these mean scores and cluster size are then modeled separately for each treatment. The presence of treatment x cluster size interaction is then tested by comparing the parameters corresponding to cluster size effects across treatments.

2. METHODOLOGY

Let $h = 1,2,\ldots,s$ index a set of cluster sizes $m_1,m_2,\ldots,m_s$; and let $i = 1,2,\ldots,t$ index treatments. Let $j = 0,1,2,\ldots,r_h$ where $r_h \leq m_h$ index the response categories which correspond to the number of subjects within clusters with the h-th cluster size who have a particular (binary response) attribute. Let π_{hij} be the probability that an experimental unit (cluster) with the h-th cluster size and the i-th treatment is classified in the j-th response category. Thus,

$$\sum_{j=0}^{r_h} \pi_{hij} = 1 \quad \text{for } \begin{matrix} h = 1,2,\ldots,s \\ i = 1,2,\ldots,t \end{matrix} . \tag{2.1}$$

As indicated in Bhapkar and Koch [1968], the analysis of the effects of treatments and cluster sizes should be investigated in terms of one or more substantively relevant functions $\phi_{hik}\ (\pi_{hi1},\ldots,\pi_{hir})$ where $k = 1,2,\ldots,u_h \leq r_h$ which suitably characterize the distribution of the response. Given the functions $\{\ \phi_{hik}\ \}$, the hypothesis of no interaction between treatments and cluster sizes can be formulated as

$$\phi_{hik} = \mu_k + \xi_{h*k} + \tau_{*ik} \quad \text{for } \begin{matrix} h = 1,2,\ldots,s \\ i = 1,2,\ldots,t \\ k = 1,2,\ldots,u_h \end{matrix} \tag{2.2}$$

where μ_k is an overall mean parameter associated with the k-th function, ξ_{h*k} is an effect due to the h-th cluster size and τ_{*ik} is an effect due to the i-th treatment and where it is understood that the $\{\xi_{h*k}\}$ and $\{\tau_{*ik}\}$ satisfy constraints such as

$$\sum_{h=1}^{s} \xi_{h*k} = 0 \text{ and } \sum_{i=1}^{t} \tau_{*ik} = 0 \quad \text{for } k = 1,2,\ldots,u_h \tag{2.3}$$

to ensure certain parameter identifiability requirements (i.e., removal of logical redundancies among parameters). For the case of additive linear

models, the ϕ_{hik} are defined by

$$\phi_{hik} = \pi_{hik} \quad \text{for } k = 1,2,\ldots,r_h, \tag{2.4}$$

while for the case of logistic linear models the ϕ_{hik} are defined by

$$\phi_{hik} = \log_e \{\pi_{hik}/\pi_{hi0}\} \quad \text{for } k = 1,2,\ldots,r_h. \tag{2.5}$$

If the data for each treatment x cluster size sub-population were not clumped so that $r_h = m_h$ for $h = 1,2,\ldots,s$, then a reasonable summary measure for the distribution of the response would be the mean rate functions

$$\phi_{hi} = \sum_{j=0}^{r_h=m_h} (j/m_h)\, \pi_{hij} \quad \text{for } \begin{matrix} h = 1,2,\ldots,s \\ i = 1,2,\ldots,t \end{matrix}. \tag{2.6}$$

The ϕ_{hi} represent the average proportion of subjects with the (binary response) attribute of interest for clusters with the h-th cluster size and the i-th treatment. Thus, the hypothesis of no interaction between treatments and cluster sizes can be formulated as indicated in (2.2) and (2.3) where the k-subscript is dropped because $u_h = 1$ for all $h = 1,2,\ldots,s$.

On the other hand, data of this type are often clumped for reasons which were outlined in Section 1. Thus, the question then arises as to whether measures analogous to the $\{\phi_{hi}\}$ in (2.6) can be formulated for these situations to reflect the average proportion of subjects with the (binary response) attribute. If the distribution of the responses within each cluster associated with the h-th cluster size and the i-th treatment could be assumed to be the standard binomial distribution with parameters (m_h,θ_{hi}) where θ_{hi} represents the probability with which subjects in such clusters exhibit the attribute, then the $\{\pi_{hij}\}$ could be expressed in terms of m_h and θ_{hi} as follows:

$$
\begin{aligned}
\pi_{hi0} &= \pi_{hi0}(\theta_{hi}) = (1 - \theta_{hi})^{m_h} \\
\pi_{hi1} &= \pi_{hi1}(\theta_{hi}) = m_h\theta_{hi}\ (1 - \theta_{hi})^{(m_h-1)} \\
&\cdots \\
\pi_{hij} &= \pi_{hij}(\theta_{hi}) = \binom{m_h}{j}\ \theta_{hi}^{j}\ (1 - \theta_{hi})^{(m_h-j)} \\
&\cdots \\
\pi_{hir_h} &= \pi_{hir_h}(\theta_{hi}) = 1 - \sum_{j=0}^{(r_h-1)} \pi_{hij}
\end{aligned}
\tag{2.7}
$$

where $r_h \leq m_h$. In this framework, the r_h-th category is the clumped category into which are classified those clusters containing at least r_h subjects with the attribute. If the model (2.7) is a valid characterization of the $\{\pi_{hij}\}$ and if $r_h = m_h$, then

$$\phi_{hi} = \theta_{hi}. \tag{2.8}$$

Thus, for such situations where the data are clumped, θ_{hi} represents a reasonable summary measure for the distribution of the response.

In order to test statistical hypotheses of no interaction like (2.2) with respect to the $\{\theta_{hi}\}$ for clumped data which satisfy the model (2.7), it is first necessary to construct functions of the observed data which represent satisfactory estimates of the $\{\theta_{hi}\}$. If n_{hi} denotes the number of clusters in the sample corresponding to the h-th cluster size and the i-th treatment and if n_{hij} denotes the number of such clusters which are classified in the j-th response category, then the likelihood function for the $((r_h+1) \times 1)$ vector $\underset{\sim}{n}_{hi}$ where $\underset{\sim}{n}'_{hi} = (n_{hi0}, n_{hi1}, \ldots, n_{hir_h})$ is

$$L_{hi}(n_{hi0}, n_{hi1}, \ldots, n_{hir_h}; \theta_{hi}) = n_{hi}! \prod_{j=0}^{r_h} \{[\pi_{hij}(\theta_{hi})]^{n_{hij}} / n_{hij}!\} \tag{2.9}$$

where the $\{\pi_{hij}(\theta_{hi})\}$ are as defined in (2.7). Thus, the maximum likelihood estimator $\hat{\theta}_{hi}$ of θ_{hi} is the solution to the equation

$$\sum_{j=0}^{r_h} \{p_{hij}/\pi_{hij}(\hat{\theta}_{hi})\}\ \{\pi^{(1)}_{hij}(\hat{\theta}_{hi})\} = 0 \qquad (2.10)$$

where $p_{hij} = (n_{hij}/n_{hi})$ is the observed proportion of clusters which are classified in the j-th response category and $\pi^{(1)}_{hij}(\hat{\theta}_{hi}) = \{(d\pi_{hij}(\theta_{hi})/d\theta_{hi})\,|\,\theta_{hi}=\hat{\theta}_{hi}\}$. Although the equation (2.10) cannot be explicitly solved to determine $\hat{\theta}_{hi}$ (unless $r_h = m_h$), it does nevertheless define $\hat{\theta}_{hi}$ implicitly as a function $F_{hi}(p_{hi0},p_{hi1},\cdots,p_{hir_h})$. Of course, the value of this function for any given observed proportion vector $\underset{\sim}{p}_{hi}$ where $\underset{\sim}{p}'_{hi} = (p_{hi0},p_{hi1},\cdots,p_{hir_h})$ must be determined by numerical methods like successive approximation search algorithms (see Kaplan and Elston [1972]). Otherwise, however, (2.10) does provide a framework which characterizes the statistical properties of $\hat{\theta}_{hi}$ as a function of $\underset{\sim}{p}_{hi}$. In particular, if the corresponding sample size n_{hi} is moderately large (e.g., $n_{hi} \geq 25$), then the statistical properties of $\hat{\theta}_{hi}$ can be investigated in an approximate sense through its stochastic linear Taylor series counterpart $f_{hi}(\underset{\sim}{p}_{hi})$ expanded about $\underset{\sim}{\pi}_{hi} = \underset{\sim}{E}\{\underset{\sim}{n}_{hi}/n_{hi}\}$ where $\underset{\sim}{\pi}'_{hi} = (\pi_{hi0},\pi_{hi1},\cdots,\pi_{hir_h})$; i.e.,

$$\hat{\theta}_{hi} = F_{hi}(\underset{\sim}{p}_{hi}) \simeq f_{hi}(\underset{\sim}{p}_{hi}) = F_{hi}(\underset{\sim}{\pi}_{hi}) + \sum_{j=0}^{r_h} [g_{hij}(\underset{\sim}{\pi}_{hi})](p_{hij}-\pi_{hij}) \qquad (2.11)$$

where $g_{hij}(\underset{\sim}{\pi}_{hi}) = \{(dF_{hi}/dp_{hij})\,|\,\underset{\sim}{p}_{hi} = \underset{\sim}{\pi}_{hi}\}$ must be determined from (2.10) by implicit differentiation techniques. In this connection, differentiation of both sides of (2.10) with respect to p_{hij} yields the equation

$$\frac{\pi^{(1)}_{hij}(F_{hi})}{\pi_{hij}(F_{hi})} + \left\{\sum_{j'=0}^{r_h} p_{hij'}\left[\frac{\{\pi_{hij'}(F_{hi})\}\ \{\pi^{(2)}_{hij'}(F_{hi})\}-\{\pi^{(1)}_{hij'}(F_{hi})\}^2}{\{\pi_{hij'}(F_{hi})\}^2}\right]\right\} g_{hij}(\underset{\sim}{p}_{hi})=0 \qquad (2.12)$$

where $F_{hi} = F_{hi}(\underset{\sim}{p}_{hi})$, $g_{hij}(\underset{\sim}{p}_{hi}) = \{dF_{hi}(\underset{\sim}{p}_{hi})/dp_{hij}\}$, $\pi^{(1)}_{hij}(F_{hi}) = \{d\pi_{hij}(F_{hi})/dF_{hi}\}$, and $\pi^{(2)}_{hij}(F_{hi}) = \{d^2\pi_{hij}(F_{hi})/dF^2_{hi}\}$. From (2.12), it follows that

$$g_{hij}(\underset{\sim}{\pi}_{hi}) = \frac{\pi^{(1)}_{hij}\{F_{hi}(\underset{\sim}{\pi}_{hi})\}/\pi_{hij}\{F_{hi}(\underset{\sim}{\pi}_{hi})\}}{\sum_{j'=0}^{r_h}\left\{\frac{[\pi^{(1)}_{hij'}\{F_{hi}(\underset{\sim}{\pi}_{hi})\}]^2}{\pi_{hij'}\{F_{hi}(\underset{\sim}{\pi}_{hi})\}} - \pi^{(2)}_{hij'}\{F_{hi}(\underset{\sim}{\pi}_{hi})\}\right\}}. \tag{2.13}$$

Since the functions $\pi_{hij}(\theta_{hi})$ in (2.7) satisfy the constraint

$$\sum_{j=0}^{r_h} \pi_{hij}(\theta_{hi}) = 1 \tag{2.14}$$

for all θ_{hi}, then the functions $\{\pi^{(1)}_{hij}(F_{hi})\}$ and $\{\pi^{(2)}_{hij}(F_{hi})\}$ necessarily satisfy the constraints

$$\sum_{j=0}^{r_h} \pi^{(1)}_{hij}(F_{hi}) = \sum_{j=0}^{r_h} \pi^{(2)}_{hij}(F_{hi}) = 0. \tag{2.15}$$

On the basis of this result, (2.13) may be simplified to

$$g_{hij}(\underset{\sim}{\pi}_{hi}) = \frac{\pi^{(1)}_{hij}\{F_{hi}(\underset{\sim}{\pi}_{hi})\}/\pi_{hij}\{F_{hi}(\underset{\sim}{\pi}_{hi})\}}{\Delta_{hi}(\underset{\sim}{\pi}_{hi})} \tag{2.16}$$

where

$$\Delta_{hi}(\underset{\sim}{\pi}_{hi}) = \sum_{j=0}^{r_h} [\pi^{(1)}_{hij}\{F_{hi}(\underset{\sim}{\pi}_{hi})\}]^2/[\pi_{hij}\{F_{hi}(\underset{\sim}{\pi}_{hi})\}] \tag{2.17}$$

Thus, from (2.11), the large sample asymptotic expectation and variance of $\hat{\theta}_{hi}$ are

$$E_A\{\hat{\theta}_{hi}\} = E\{f_{hi}(\underset{\sim}{p}_{hi})\} = F_{hi}(\underset{\sim}{\pi}_{hi}) \tag{2.18}$$

$$\mathrm{Var}_A\{\hat{\theta}_{hi}\} = \mathrm{Var}\{f_{hi}(\underset{\sim}{p}_{hi})\}$$
$$= \frac{1}{n_{hi}}\left\{\left[\sum_{j=0}^{r_h}\{g_{hij}(\underset{\sim}{\pi}_{hi})\}^2\pi_{hij}\right] - \left[\sum_{j=0}^{r_h}\{g_{hij}(\underset{\sim}{\pi}_{hi})\}\pi_{hij}\right]^2\right\} \tag{2.19}$$

where "E_A" denotes asymptotic expectation and "Var_A" denotes asymptotic variance. If the $\{\pi_{hij}\}$ do in fact satisfy (2.7), then (2.10) implies that

$$E_A\{\hat{\theta}_{hi}\} = F_{hi}(\underset{\sim}{\pi}_{hi}) = \theta_{hi} \tag{2.20}$$

and (2.15) - (2.17) imply that

$$\mathrm{Var}_A\{\hat{\theta}_{hi}\} = \{1/[n_{hi}\ \overline{\Delta}_{hi}(\theta_{hi})]\} \tag{2.21}$$

where

$$\overline{\Delta}_{hi}(\theta_{hi}) = \sum_{j=0}^{r_h}[\pi_{hij}^{(1)}(\theta_{hi})]^2/[\pi_{hij}(\theta_{hi})]\ . \tag{2.22}$$

The result in expression (2.21) can be alternatively derived as the negative inverse of the Fisher Information Matrix since

$$-E\left\{\frac{d^2\ \log_e L_{hi}}{d\theta_{hi}^2}\right\} = n_{hi}\ \overline{\Delta}_{hi}(\theta_{hi}). \tag{2.23}$$

Consistent estimators for the variance of $\hat{\theta}_{hi}$ may be developed from two points of view. In the first case, (2.10) is used to define θ_{hi} as an implicit function $F_{hi}(\underset{\sim}{\pi}_{hi})$ of $\underset{\sim}{\pi}_{hi}$ in the same sense that (2.6) defines ϕ_{hi} as an explicit function of $\underset{\sim}{\pi}_{hi}$. Since (2.8) demonstrates that ϕ_{hi} and θ_{hi} are equal under the model (2.7) whenever $r_h = m_h$, it follows that θ_{hi} can be interpreted in general (regardless of the validity of (2.7)) as a clumped data analogue of ϕ_{hi} in the sense of being a measure of the average proportion of subjects with the (binary response) attribute of interest for clusters

with the h-th cluster size and the i-th treatment. However, the statistic $\hat{\theta}_{hi}$ is the corresponding sample estimator function $F_{hi}(\underset{\sim}{p}_{hi})$ based on $\underset{\sim}{p}_{hi}$. For this reason, if π_{hij} is replaced by p_{hij} in (2.19), the the resulting statistic

$$V_{hi} = V_{hi}(\underset{\sim}{p}_{hi}) = \frac{1}{n_{hi}} \left\{ \left[\sum_{j=0}^{r_h} \{g_{hij}(\underset{\sim}{p}_{hi})\}^2 p_{hij} - \sum_{j=0}^{r_h} \{g_{hij}(\underset{\sim}{p}_{hi})\} \underset{\sim}{p}_{hi} \right]^2 \right\} \quad (2.24)$$

is a consistent estimator for the variance of $\hat{\theta}_{hi}$ which does not require the validity of (2.7). Moreover, in view of (2.10), (2.24) may be simplified to

$$V_{hi} = V_{hi}(\underset{\sim}{p}_{hi}) = \frac{1}{n_{hi}} \left\{ \sum_{j=0}^{r_h} \{g_{hij}(\underset{\sim}{p}_{hi})\}^2 \, p_{hij} \right\} , \quad (2.25)$$

where

$$g_{hij}(\underset{\sim}{p}_{hi}) = \{\pi^{(1)}_{hij}(\hat{\theta}_{hi})/\pi_{hij}(\hat{\theta}_{hi})\} \left\{ \sum_{j'=0}^{r_h} [\pi^{(1)}_{hij'}(\hat{\theta}_{hi})]^2/[\pi_{hij}(\hat{\theta}_{hi})] \right\}^{-1} \quad (2.26)$$

On the other hand, if the $\{\pi_{hij}\}$ do in fact satisfy the model (2.7), then an alternative estimator for the variance of $\hat{\theta}_{hi}$ can be obtained by replacing θ_{hi} by $\hat{\theta}_{hi}$ in (2.21). The resulting statistic is

$$\overline{V}_{hi} = \overline{V}_{hi}(\hat{\theta}_{hi}) = \{1/[n_{hi} \, \overline{\Delta}_{hi}(\hat{\theta}_{hi})]\}. \quad (2.27)$$

Since $\overline{V}_{hi}$ is based entirely on $\hat{\theta}_{hi}$, it has somewhat more stable statistical properties than V_{hi} which involves both $\hat{\theta}_{hi}$ and $\underset{\sim}{p}_{hi}$ when the model (2.7) holds; it also is easier to compute. For this reason, it is preferable to use $\overline{V}_{hi}$ in those cases for which the validity of (2.7) is a reasonable assumption.

Thus, it is of interest to test (2.7) as a hypothesis on the vector $\underset{\sim}{\pi}_{hi}$.

One appropriate statistic which can be used for this purpose is the Pearson χ^2-statistic

$$X^2_{P,hi} = \sum_{j=0}^{r_h} \frac{\{n_{hij} - n_{hi}\pi_{hij}(\hat{\theta}_{hi})\}^2}{n_{hi}\pi_{hij}(\hat{\theta}_{hi})} \tag{2.28}$$

which has approximately the χ^2-distribution with D.F. = $(r_h - 2)$ when (2.7) holds. If $X^2_{P,hi}$ is small in the sense of not being significant at the $\alpha = 0.25$ level, then it is appropriate to use $\overline{V}_{hi}$ as the estimator of the variance of $\hat{\theta}_{hi}$; otherwise, V_{hi} should be used. However, if several treatment x cluster size sub-populations are being considered simultaneously, this rule should be relaxed in the sense of being applied only to

$$X^2_P = \sum_{h=1}^{s} \sum_{i=1}^{t} X^2_{P,hi} \tag{2.29}$$

instead of the individual $\{X^2_{P,hi}\}$; the statistic X^2_P has approximately the χ^2-distribution with D.F. = $t\left\{\sum_{h=1}^{s} (r_h - 2)\right\}$ when (2.7) holds for all combinations of h = 1,2,...,s and i = 1,2,...,t.

In summary, the parameters $\{\theta_{hi}\}$ are functions $\{F_{hi}(\underset{\sim}{\pi}_{hi})\}$ which are measures of "location" (or central tendency) for the distributions of response (i.e., the numbers of subjects with a particular binary attribute) for clusters from the respective treatment x cluster size sub-populations. For this purpose, they can be interpreted as clumped data analogues of the mean score functions $\{\phi_{hi}\}$ in (2.6) in the sense that both tend to increase as the corresponding distributions become more concentrated in the response categories with higher occurrence rates for the binary attribute (i.e., larger j) and less concentrated in the response categories associated with lower occurrence rates (i.e., smaller j). Thus, they provide a reasonable criterion in terms

of which the effects of treatments and cluster sizes can be investigated with respect to differences (shift alternatives) in "location." In this context, the hypothesis of no interaction between treatments and cluster sizes can be formulated for the $\{\theta_{hi}\}$ in the manner indicated in (2.2) and (2.3) as

$$\theta_{hi} = \mu + \xi_{h*} + \tau_{*i} \tag{2.30}$$

where μ is an overall mean parameter, ξ_{h*} is an effect due to the h-th cluster size, τ_{*i} is an effect due to the i-th treatment, and where it is usually understood that the $\{\xi_{h*}\}$ and $\{\tau_{*i}\}$ satisfy constraints such as

$$\sum_{h=1}^{s} \xi_{h*} = 0, \quad \sum_{i=1}^{t} \tau_{*i} = 0 \tag{2.31}$$

to ensure certain parameter identifiability requirements (i.e., removal of logical redundancies among parameters).

Test statistics for hypotheses like (2.30) and estimators for corresponding model parameters can be obtained by using the general approach for the analysis of multivariate categorical data described by Grizzle, Starmer, and Koch [1969] (henceforth abbreviated GSK) as extended to implicit functions by Koch and Tolley [1975]. For this purpose, the $\{\theta_{hi}\}$ are regarded as implicit functions $\{F_{hi}(\underset{\sim}{\pi}_{hi})\}$ of the underlying probability vectors $\{\underset{\sim}{\pi}_{hi}\}$; and their respective estimators $\{\hat{\theta}_{hi}\}$ are the corresponding functions of the observed proportion vectors $\{\underset{\sim}{p}_{hi}\}$. The estimators $\{V_{hi}\}$ in (2.25) for the variances of the $\{\theta_{hi}\}$ are essentially the same as those given in GSK except that they require the use of first partial derivatives which are obtained via implicit function techniques. Similarly, the $\{\overline{V}_{hi}\}$ represent somewhat more refined estimators of variance which are applicable when the model (2.7) is valid. Thus, when the sample size $\{n_{hi}\}$ are sufficiently large (e.g., $\min\{n_{hi}\} \geq 25$), the statistics $\{\hat{\theta}_{hi}\}$ can be analyzed by the same types of

Wald [1943] statistics and the corresponding weighted least squres computational algorithms described in Bhapkar and Koch [1968a, 1968b] and GSK. Here, however, the estimators $\{\hat{\theta}_{hi}\}$ and their estimated variances $\{V_{hi}\}$ (or $\{\overline{V}_{hi}\}$) must be determined by suitably specialized computer sub-routines for successive approximation and direct caluclation as opposed to matrix multiplication. Otherwise, a more detailed discussion of this phase of the methodology is presented with respect to specific models of interest for the example in Section 3.

Finally, it should be recognized that the $\{\theta_{hi}\}$ only pertain to the "location" (or central tendency) of the response distribution. In many situations, such parameters may be the only ones of practical interest; and hence no loss of generality occurs when the analysis is restricted to them. Moreover, this type of approach is entirely appropriate if the clumped binomial model (2.7) is valid, in which case, the $\{\theta_{hi}\}$ provide complete information about the nature of the respective distributions. On the other hand, if the model (2.7) does not apply (e.g., because its validity is contradicted by the test statistics (2.28) or (2.29)) and the analysis of "shape" parameters is also viewed as practically important, then analogous methods used on clumped discrete distributions other than the binomial are required. Further discussion of this topic is given in Tolley and Koch [1974] for certain contagious distributions.

3. EXAMPLE

This section is concerned with the analysis of the animal depletion data in Table 1 which was originally reported by Kastenbaum and Lamphiear [1959]. Primary emphasis will be given to illustrating the methodology in Section 2. The results obtained from this analysis are then compared with those based on certain related alternative methods.

3.1. Weighted least squares analysis of the mean rate functions $\{\theta_{hi}\}$

For the data in Table 1, there are s = 5 litter size types which are specifically 7, 8, 9, 10, 11; there are t = 2 treatments called A and B. The response of interest is the number of deaths in litters of baby mice prior to weaning and the corresponding distribution involves r = 3 categories which are "0 depletions," "1 depletion," and the clumped category of "2 or more depletions."

The estimated mean rate parameters $\{\hat{\theta}_{hi}\}$ defined by (2.10) are given in Table 2 together with corresponding estimated standard errors based on the $\{V_{hi}\}$ in (2.25) and the $\{\overline{V}_{hi}\}$ in (2.27). In addition, Table 2 contains

TABLE 2

OBSERVED AND PREDICTED VALUES FOR CLUMPED BINOMIAL MEAN RATE FUNCTIONS, STANDARD ERRORS, AND GOODNESS OF FIT STATISTICS

Litter size	Treatment	Observed estimates $\hat{\theta}_{hi}$	Estimated s.e.$(\hat{\theta}_{hi})$		Final model fit based on $\underline{V}$		Final model fit based on $\overline{\underline{V}}$		Binomial model goodness of fit
			Unrestricted $(V_{hi})^{\frac{1}{2}}$	Restricted $(\overline{V}_{hi})^{\frac{1}{2}}$	Predicted values	Estimated s.e.	Predicted values	Estimated s.e.	$X^2_{P,hi}$(D.F.=1)
7	A	0.041	0.010	0.009	0.040	0.006	0.041	0.005	5.21
7	B	0.048	0.009	0.008	0.040	0.006	0.041	0.005	3.71
8	A	0.060	0.011	0.010	0.066	0.005	0.066	0.004	6.83
8	B	0.051	0.010	0.009	0.054	0.004	0.054	0.004	3.93
9	A	0.087	0.013	0.012	0.091	0.007	0.091	0.006	3.24
9	B	0.063	0.010	0.009	0.068	0.005	0.067	0.005	1.11
10	A	0.113	0.017	0.016	0.117	0.010	0.115	0.009	1.23
10	B	0.079	0.011	0.010	0.082	0.007	0.081	0.007	3.37
11	A	0.157	0.021	0.022	0.142	0.013	0.140	0.013	0.51
11	B	0.112	0.017	0.019	0.096	0.010	0.094	0.010	3.15

the goodness of fit statistics $\{X^2_{P,hi}\}$ defined by (2.28) for testing the validity of the underlying binomial models (2.7). Since the sum of these test statistics $X^2_P = 32.29$ with D.F. = 10 is significant ($\alpha = .01$), it follows that these data are not satisfactorily characterized by the binomial model (2.7) in a strict sense.

For this reason, the $\{V_{hi}\}$ must be used as the estimators of variance for the $\{\hat{\theta}_{hi}\}$ because the assumption underlying the validity of the $\{\overline{V}_{hi}\}$ does not appear to hold. Nevertheless, results associated with further analyses of the $\{\hat{\theta}_{hi}\}$ will be presented in parallel for both the case where the $\{V_{hi}\}$ are used as estimators of variance as well as the case where the $\{\overline{V}_{hi}\}$ are so used for purposes of comparison. However, in this context, it should be recognized that those based on the $\{V_{hi}\}$ have reasonable validity while those based on the $\{\overline{V}_{hi}\}$ do not.

To apply the GSK approach to analyze the effects of treatment and litter size on the $\{\hat{\theta}_{hi}\}$, let $\underset{\sim}{F}$ in (3.1) denote the vector of estimators $\{\hat{\theta}_{hi}\}$, and let $\underset{\sim}{V}$ and $\underset{\sim}{\overline{V}}$ denote the corresponding estimated covariance matrices based on the $\{V_{hi}\}$ and the $\{\overline{V}_{hi}\}$ respectively.

$$\underset{\sim}{F} = \begin{bmatrix} \hat{\theta}_{11} \\ \hat{\theta}_{12} \\ \hat{\theta}_{21} \\ \hat{\theta}_{22} \\ \hat{\theta}_{31} \\ \hat{\theta}_{32} \\ \hat{\theta}_{41} \\ \hat{\theta}_{42} \\ \hat{\theta}_{51} \\ \hat{\theta}_{52} \end{bmatrix}, \quad \underset{\sim}{V} = \begin{bmatrix} V_{11} & & & & & & & & & \\ & V_{12} & & & & & & & & \\ & & V_{21} & & & & & & & \\ & & & V_{22} & & & & & & \\ & & & & V_{31} & & & & & \\ & & & & & V_{32} & & & & \\ & & & & & & V_{41} & & & \\ & & & & & & & V_{42} & & \\ & & & & & & & & V_{51} & \\ & & & & & & & & & V_{52} \end{bmatrix}, \quad \underset{\sim}{\overline{V}} = \begin{bmatrix} \overline{V}_{11} & & & & & & & & & \\ & \overline{V}_{12} & & & & & & & & \\ & & \overline{V}_{21} & & & & & & & \\ & & & \overline{V}_{22} & & & & & & \\ & & & & \overline{V}_{31} & & & & & \\ & & & & & \overline{V}_{32} & & & & \\ & & & & & & \overline{V}_{41} & & & \\ & & & & & & & \overline{V}_{42} & & \\ & & & & & & & & \overline{V}_{51} & \\ & & & & & & & & & \overline{V}_{52} \end{bmatrix} \tag{3.1}$$

Preliminary analysis of $\underset{\sim}{F}$ was undertaken in terms of the complete factorial model

$$\underset{\sim}{E_A}\{\underset{\sim}{F}\} = \underset{\sim}{X_1}\underset{\sim}{\beta_1} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & -1 & 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & -1 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & -1 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 \\ 1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \\ 1 & -1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \tau_{*1} \\ \xi_{1*} \\ \xi_{2*} \\ \xi_{3*} \\ \xi_{4*} \\ (\xi\tau)_{11} \\ (\xi\tau)_{21} \\ (\xi\tau)_{31} \\ (\xi\tau)_{41} \end{bmatrix}, \qquad (3.2)$$

where "$\underset{\sim}{E_A}$" denotes asymptotic expectation, μ is an overall mean parameter, τ_{*1} is an effect due to treatment A, the $\{\xi_{h*}\}$ are effects due to litter size for h = 1,2,3,4, and the $\{(\xi\tau)_{h1}\}$ are effects due to treatment x litter size interaction for h = 1,2,3,4.

The model $\underset{\sim}{X_1}$ as well as the others presented later in this paper were fitted by the method of weighted least squares as described in general by GSK; and in the specific context of implicitly defined statistics by Koch and Tolley [1975]. Statistical tests X^2_W which are based on this methodology for hypotheses involving the parameters in $\underset{\sim}{\beta_1}$ are shown in Table 3. The

TABLE 3

TEST STATISTICS FOR COMPLETE $\underset{\sim}{X_1}$ MODEL

Source of variation	D.F.	X^2_W-tests based on $\underset{\sim}{V}$	X^2_W-tests based on $\underset{\sim}{\overline{V}}$
Treatment	1	2.15	2.49
Litter size	4	47.30	48.29
Treatment x litter size	4	5.21	5.79
Non-linear litter size	3	2.47	2.52
Treatment x non-linear litter size	3	0.06	0.06

principal conclusion that emerges from these results is that the non-linear components of both litter size effects and treatment x litter size interaction are non-significant ($\alpha = 0.25$) and can be eliminated from further consideration.

Thus, attention was then directed at the revised model

$$\underset{\sim}{E}_{\underset{\sim}{A}}\{\underset{\sim}{F}\} = \underset{\sim}{X}_2\underset{\sim}{\beta}_2 = \begin{bmatrix} 1 & 1 & -2 & -2 \\ 1 & -1 & -2 & 2 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & 2 & 2 \\ 1 & -1 & 2 & -2 \end{bmatrix} \begin{bmatrix} \mu \\ \tau_{*1} \\ \lambda \\ (\lambda\tau)_{*1} \end{bmatrix}, \qquad (3.3)$$

where μ is an overall mean parameter, τ_{*1} is an effect due to treatment A, λ is an effect due to the linear component of litter size, and $(\lambda\tau)_{*1}$ is an effect due to the interaction of treatment with the linear component of litter size. Statistical tests of hypotheses involving these parameters are shown in Table 4. In particular, these results indicate significant ($\alpha = .05$) interaction between treatment and litter size in the sense of its

TABLE 4

TEST STATISTICS FOR REDUCED $\underset{\sim}{X}_2$ MODEL

Source of Variation	D.F.	X^2_W-tests based on $\underset{\sim}{V}$	X^2_W-tests based on $\overline{\underset{\sim}{V}}$
Treatment	1	6.96	7.91
Linear litter size	1	45.24	46.62
Treatment x linear litter size	1	5.20	5.98
Overall model	3	47.36	49.08
Residual	6	2.81	2.74

linear component; i.e., as the litter size becomes larger the difference between $\hat{\theta}_{h1}$ and $\hat{\theta}_{h2}$ becomes larger which means that the difference between treatments A and B in the rate of depletions increases with litter size.

Further examination of the predicted values for the $\{\theta_{hi}\}$ derived from the model $\underset{\sim}{X}_2$ suggested that there was essentially no real difference between treatments A and B for litters of size 7. Thus, it was concluded that the final model shown in (3.4) suitably characterized the effects of treatment and litter size on the $\{\theta_{hi}\}$.

$$\underset{\sim}{E}_{\underset{\sim}{A}}\{\underset{\sim}{F}\} = \underset{\sim}{X}_3\underset{\sim}{\beta}_3 = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 2 & 0 \\ 1 & 0 & 2 \\ 1 & 3 & 0 \\ 1 & 0 & 3 \\ 1 & 4 & 0 \\ 1 & 0 & 4 \end{bmatrix} \begin{bmatrix} \eta \\ \lambda_1 \\ \lambda_2 \end{bmatrix} \tag{3.4}$$

Here, η is an intercept parameter corresponding to the common predicted value for both treatments for litters of size 7, λ_1 is an effect due to the linear component of litter size for treatment A, and λ_2 is an effect due to the linear component of litter size for treatment B. Estimates for these parameters and their corresponding standard errors are given in Table 5; statis-

TABLE 5

ESTIMATED PARAMETERS AND STANDARD ERRORS FOR FINAL $\underset{\sim}{X}_3$ MODEL

Parameter	Final model fit based on $\underset{\sim}{V}$		Final model fit based on $\underset{\sim}{\bar{V}}$	
	Estimated parameter	Estimated s.e.	Estimated parameter	Estimated s.e.
η	0.040	0.006	0.041	0.005
λ_1	0.025	0.004	0.025	0.004
λ_2	0.014	0.003	0.013	0.003

tical tests of hypotheses are given in Table 6. Predicted values for the

TABLE 6

TEST STATISTICS FOR FINAL $\underset{\sim}{X}_3$ MODEL

Source of variation	D.F.	X^2_W-tests based on $\underset{\sim}{V}$	X^2_W-tests based on $\overline{\underset{\sim}{V}}$
Linear litter size for treatment A	1	42.71	44.54
Linear litter size for treatment B	1	19.68	19.24
Treatment x linear litter size	1	8.39	9.10
Overall model	2	47.16	48.80
Residual	7	3.01	3.03

$\{\theta_{hi}\}$ based on the final model $\underset{\sim}{X}_3$ are given in Table 2 with corresponding standard errors. From these latter results, it is apparent that these predicted values both provide a good fit to the original observed estimates $\{\hat{\theta}_{hi}\}$ as well as have substantially smaller estimated standard errors.

In summary, if the data in Table 1 are analyzed with respect to the mean rate functions $\{\theta_{hi}\}$, then a significant interaction between treatment and litter size can be detected. Moreover, the nature of this interaction is such that there is essentially no difference between treatments A and B for litters of size 7, but a linearly increasing difference between the rates for treatments A and B for the larger litters of sizes 8,9,10,11. These findings are graphically displayed in Figure 1. Similarly, Figure 2 represents a corresponding graph for the predicted values for the quantities $\{m_h\theta_{hi}\}$ which can be interpreted as the implied expected number of deple-

FIGURE 1

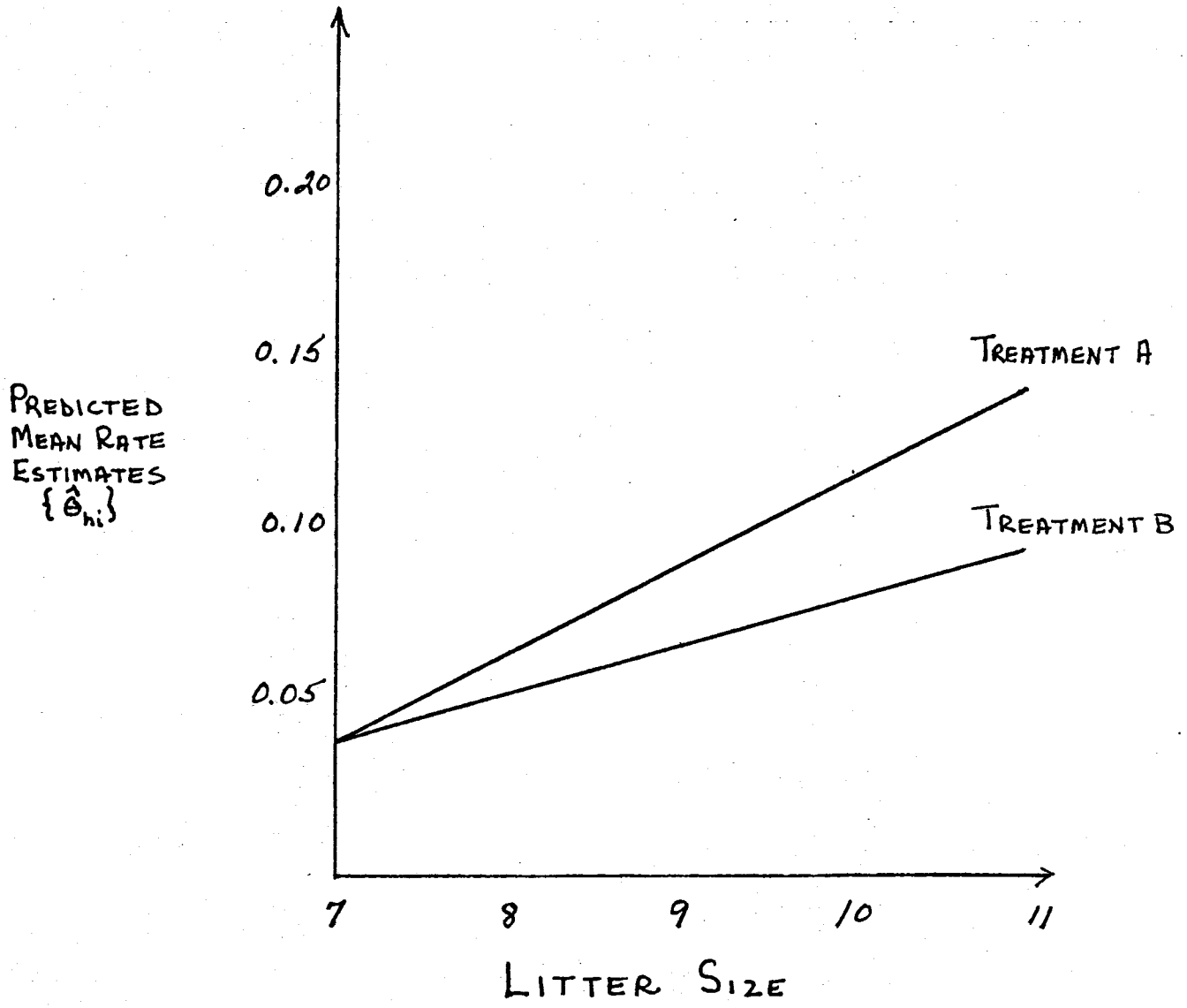
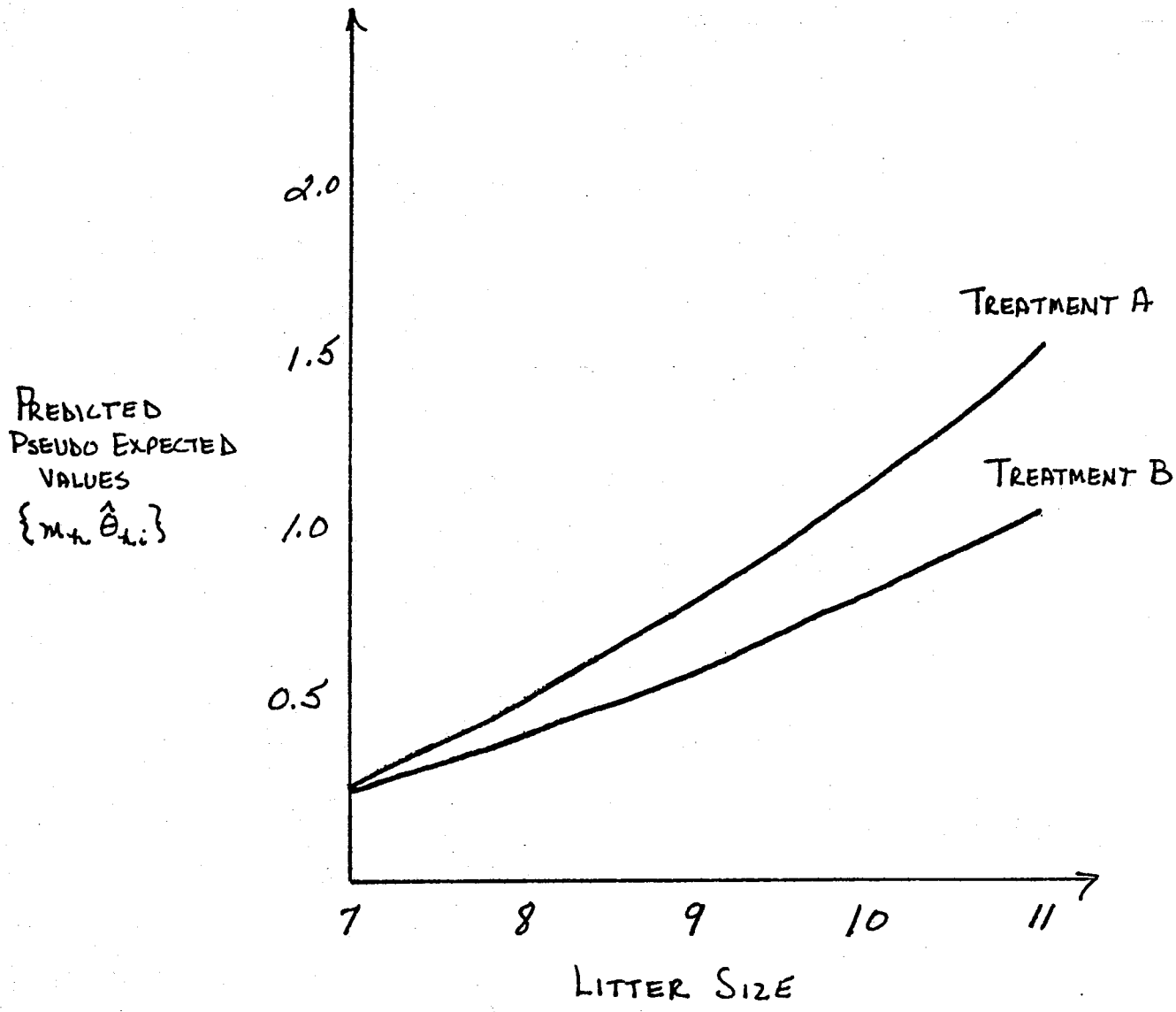

PREDICTED MEAN RATE ESTIMATES $\{\hat{\theta}_{hi}\}$
0.20
0.15
0.10
0.05
TREATMENT A
TREATMENT B
7
8
9
10
11
LITTER SIZE

FIGURE 2


PREDICTED
PSEUDO EXPECTED
VALUES
$\{m_t \hat{\theta}_{t.}\}$
2.0
1.5
1.0
0.5
TREATMENT A
TREATMENT B
7
8
9
10
11
LITTER SIZE

tions for the underlying fitted clumped binomial model. Since it has been previously noted that the data in Table 1 are not strictly characterized by the clumped binomial model (2.7), the $\{m_h\ \theta_{hi}\}$ should be viewed with some caution in the sense that they are pseudo expected values as opposed to true ones.

Finally, although statistical analysis can be directed at the parameters $\{\theta_{hi}\}$ regardless of the validity of the model (2.7), the interpretation of such results should be guided by considerations which reflect the effectiveness of this type of approach. For this reason, it is useful to formulate "measures of explained variation" which can be associated with certain models. One such measure is given by

$$R_M^2 = 1 - (X_{P,M}^2/X_{P,U}^2) \tag{3.5}$$

where $X_{P,M}^2$ is the Pearson χ^2-statistic for goodness of fit of the model of interest and $X_{P,U}^2$ is the Pearson χ^2-statistic for goodness of fit of the model for total uniformity in the observed contingency table (i.e., $\pi_{hi0} = \pi_{hi1} = \pi_{hi2} = (1/3)$ for h = 1,2,3,4,5 and i = 1,2). For the data in Table 1, $X_{P,U}^2$ =291.94; and thus $R_{M1}^2 = 0.89$ for the clumped binomial model (2.7). Similarly, for the final model corresponding to (2.7) in combination with (3.4), $R_{M3}^2 = 0.88$. From these results, it follows that the clumped binomial model provides a very effective framework for analyzing the variation in the data in Table 1.

3.2. A bivariate clumped function analysis

The major disadvantage associated with the analysis presented in Section 3.1 is the potentially difficult successive approximation computations required to determine the estimators $\{\hat{\theta}_{hi}\}$. For this reason, it is of interest to consider an alternative approach which is based on explicit func-

tions which are meaningful with respect to the clumped binomial model (2.7). In particular, the quantities $\hat{\theta}_{hi1}$ and $\hat{\theta}_{hi2}$ in (3.6)

$$\begin{aligned} \hat{\theta}_{hi1} &= 1 - \exp\{(1/m_h)\ [\log_e(p_{hi0})]\} \\ \hat{\theta}_{hi2} &= \{1 + \exp[\log_e(m_h p_{hi0}/p_{hi1})]\}^{-1} \end{aligned} \tag{3.6}$$

are compound function estimators of θ_{hi} if the model (2.7) is valid. These statistics together with corresponding estimated standard errors and estimated covariance (as determined by the methods described in Forthofer and Koch [1973]) are given in Table 7. In addition, Table 7 contains weighted

TABLE 7

ESTIMATED PARAMETERS AND STANDARD ERRORS FOR BIVARIATE FUNCTION ANALYSIS

litter Size	Treatment	$\hat{\theta}_{hi1}$	Estimated s.e.$(\hat{\theta}_{hi1})$	$\hat{\theta}_{hi2}$	Estimated s.e.$(\hat{\theta}_{hi2})$	Estimated Cov$(\hat{\theta}_{hi1},\hat{\theta}_{hi2})$	$X^2_{W,hi}$(D.F.=1) for $\hat{\theta}_{hi1}$ vs $\hat{\theta}_{hi2}$	$\hat{\theta}_{hi\cdot}$	Estimated s.e.$(\hat{\theta}_{hi\cdot})$
7	A	0.034	0.008	0.026	0.008	0.000061	3.05	0.030	0.008
7	B	0.042	0.008	0.035	0.009	0.000062	2.78	0.040	0.008
8	A	0.049	0.010	0.034	0.010	0.000081	5.69	0.043	0.009
8	B	0.044	0.009	0.035	0.009	0.000070	3.25	0.042	0.009
9	A	0.074	0.013	0.057	0.016	0.000168	3.91	0.076	0.013
9	B	0.058	0.010	0.052	0.013	0.000114	1.16	0.060	0.010
10	A	0.100	0.019	0.080	0.028	0.000440	1.67	0.107	0.018
10	B	0.068	0.011	0.053	0.013	0.000121	4.12	0.070	0.011
11	A	0.175	0.035	0.214	0.097	0.003158	0.36	0.157	0.020
11	B	0.145	0.032	0.214	0.087	0.002617	1.45	0.111	0.014

least squares $\{X^2_{W,hi}\}$ statistics for testing the validity of the model (2.7) in terms of the implied hypotheses

$$E_A\{\hat{\theta}_{hi1} - \hat{\theta}_{hi2}\} = 0 \quad , \tag{3.7}$$

as well as the estimators $\{\hat{\theta}_{hi\cdot}\}$ which are based on the weighted least squares pooling of the $\{\hat{\theta}_{hi1}\}$ and $\{\hat{\theta}_{hi2}\}$ under a model which presumes their respective equality. Thus, the $\hat{\theta}_{hi\cdot}$ are analogous to the $\{\hat{\theta}_{hi}\}$ in the sense that both represent combined estimators based on all the data for each of the respective treatment x litter size sub-populations.

Tests statistics associated with the analyses of the $\{\hat{\theta}_{hi1}\}$ and the $\{\hat{\theta}_{hi2}\}$ in a combined bivariate framework for the model $\underset{\sim}{X}_1$ in (3.2) are given in Table 8 together with those corresponding to the $\{\hat{\theta}_{hi\cdot}\}$ in a uni-

TABLE 8

TEST STATISTICS FOR BIVARIATE COMPLETE $\underset{\sim}{X}_1$ MODEL

Source of variation	D.F. for univariate tests	X^2_W-tests for $\hat{\theta}_{hi1}$	X^2_W-tests for $\hat{\theta}_{hi2}$	X^2_W-tests for $\hat{\theta}_{hi\cdot}$	D.F. for bivariate tests	X^2_W-tests for $(\hat{\theta}_{hi1}, \hat{\theta}_{hi2})$
Treatment	1	1.00	0.08	0.96	2	2.30
Litter size	4	39.11	14.96	71.04	8	60.99
Treatment x litter size	4	3.42	1.28	7.46	8	5.78
Non-linear litter size	3	6.78	5.20	5.68	6	7.26
Treatment x non-linear litter size	3	0.08	0.15	0.12	6	0.24
Total	9	40.10	15.70	72.35	18	64.80

variate context. Similar results for the reduced model $\underset{\sim}{X}_2$ in (3.3) are shown in Table 9. Thus, a significant ($\alpha = .05$) interaction between treatment and the linear component of litter size is detected for the $\{\hat{\theta}_{hi1}\}$ and the $\{\hat{\theta}_{hi\cdot}\}$. However, neither the treatment effect nor its interaction with litter size are significant ($\alpha = .25$) for the $\{\hat{\theta}_{hi2}\}$.

TABLE 9

TEST STATISTICS FOR BIVARIATE REDUCED $\underset{\sim}{X}_2$ MODEL

Source of variation	D.F. for univariate tests	X^2_W-tests for $\hat{\theta}_{hi1}$	X^2_W-tests for $\hat{\theta}_{hi2}$	X^2_W-tests for $\hat{\theta}_{hi\cdot}$	D.F. for bivariate tests	X^2_W-tests for $(\hat{\theta}_{hi1}, \hat{\theta}_{hi2})$
Treatment	1	4.56	0.67	5.07	2	7.31
Linear litter size	1	46.98	16.91	65.71	2	54.05
Treatment x linear litter size	1	4.49	1.32	6.72	2	5.57
Overall model	3	47.24	17.43	66.08	6	56.53
Residual	6	xxxx	xxxx	6.27	12	8.28

From these considerations, it follows that the model $\underset{\sim}{X}_3$ in (3.4) also suitably characterizes the effects of treatment and litter size on the $\{\hat{\theta}_{hi\cdot}\}$. Estimated parameters based on this latter analysis are given in Table 10; and

TABLE 10

ESTIMATED PARAMETERS AND STANDARD ERRORS FOR POOLED BIVARIATE FINAL $\underset{\sim}{X}_3$ MODEL

Parameter	Estimated parameter	Estimated s.e.
η	0.029	0.005
λ_1	0.026	0.004
λ_2	0.017	0.003

statistical tests of hypotheses are given in Table 11. Thus, the conclu-

TABLE 11

TEST STATISTICS FOR POOLED BIVARIATE FINAL $\underset{\sim}{X}_3$ MODEL

Source of variation	D.F.	X^2_W
Linear litter size for treatment A	1	51.84
Linear litter size for treatment B	1	34.75
Treatment x linear litter size	1	6.84
Overall model	2	64.96
Residual	7	7.39

sions for the analysis of the $\{\hat{\theta}_{hi\cdot}\}$ are essentially the same as those for the analysis of the $\{\hat{\theta}_{hi}\}$ in Section 3.1. These same remarks also apply to the joint analysis of the $\hat{\theta}_{hi1}$ and $\hat{\theta}_{hi2}$ as reflected by the final model $\underset{\sim}{X}_4$

$$\underset{\sim}{E}_A \begin{bmatrix} \hat{\theta}_{111} \\ \hat{\theta}_{112} \\ \hat{\theta}_{121} \\ \hat{\theta}_{122} \\ \hat{\theta}_{211} \\ \hat{\theta}_{212} \\ \hat{\theta}_{221} \\ \hat{\theta}_{222} \\ \hat{\theta}_{311} \\ \hat{\theta}_{312} \\ \hat{\theta}_{321} \\ \hat{\theta}_{322} \\ \hat{\theta}_{411} \\ \hat{\theta}_{412} \\ \hat{\theta}_{421} \\ \hat{\theta}_{422} \\ \hat{\theta}_{511} \\ \hat{\theta}_{512} \\ \hat{\theta}_{521} \\ \hat{\theta}_{522} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 2 & 0 \\ 1 & 0 & 2 \\ 1 & 0 & 2 \\ 1 & 0 & 2 \\ 1 & 3 & 0 \\ 1 & 0 & 3 \\ 1 & 0 & 3 \\ 1 & 0 & 3 \\ 1 & 4 & 0 \\ 1 & 0 & 4 \\ 1 & 0 & 4 \\ 1 & 0 & 4 \end{bmatrix} \begin{bmatrix} \eta \\ \gamma_1 \\ \gamma_2 \end{bmatrix} = \underset{\sim}{X}_4 \underset{\sim}{\beta}_4 \quad , \qquad (3.8)$$

where η is an intercept parameter corresponding to the common predicted value for both treatments and both functions for litters of size 7, γ_1 is an effect due to the linear component of litter size for treatment A for the $\{\hat{\theta}_{hi1}\}$, and γ_2 is an effect due to the linear component of litter size for treatment B for the $\{\hat{\theta}_{hi1}\}$ and both treatments for the $\{\hat{\theta}_{hi2}\}$. Estimators for these parameters and their corresponding standard errors are shown in Table 12; and statistical tests of hypotheses are shown in Table 13.

TABLE 12

ESTIMATED PARAMETERS AND STANDARD ERRORS FOR BIVARIATE FINAL $\underset{\sim}{X}_4$ MODEL

Parameter	Estimated parameter	Estimated s.e.
η	0.030	0.005
γ_1	0.024	0.003
γ_2	0.016	0.003

TABLE 13

TEST STATISTICS FOR BIVARIATE FINAL $\underset{\sim}{X}_4$ MODEL

Source of variation	D.F.	X^2_W
Linear litter size for treatment A for $\{\hat{\theta}_{hi1}\}$	1	75.74
Linear litter size for treatment B for $\{\hat{\theta}_{hi1}\}$ and for both treatments for $\{\hat{\theta}_{hi2}\}$	1	33.77
Treatment x linear litter size interaction	1	18.20
Overall model	2	76.32
Residual	17	23.47

Thus, it follows that treatment effects and the interaction of treatment with the linear component of litter size primarily pertain to the $\{\hat{\theta}_{hi1}\}$ which are associated with the probabilities of "0" depletions. Finally, the fact that the model $\underset{\sim}{X}_4$ implies somewhat different conclusions for the $\{\hat{\theta}_{hi1}\}$ and $\{\hat{\theta}_{hi2}\}$ partially accounts for the limitations of the clumped binomial model (2.7) for providing a strictly valid characterization of the data in Table 1 in the sense of the goodness of fit statistics in (2.28) and (2.29) or those associated with (3.7).

3.3. Other analyses

The methods given in Sections 3.1 and 3.2 are relatively complex. For this reason, the question then arises as to whether the results which are obtained from them are consistent with those suggested by simpler and more well-known procedures. Thus, it is useful to note that the overall Pearson χ^2-statistic for the data in Table 1 is $X_P^2 = 107.01$ with D.F. = 18 which indicates that significant differences ($\alpha = .01$) exist among the distributions for the respective treatment x litter size sub-populations. Similarly, Pearson χ^2-statistics for litter size within each treatment are

$$\begin{aligned} X_P^2 &= 57.33 \quad \text{with D.F.} = 8 \quad \text{for treatment A} \\ X_P^2 &= 39.56 \quad \text{with D.F.} = 8 \quad \text{for treatment B;} \end{aligned} \tag{3.9}$$

and for treatment within each litter size are

$$\begin{aligned} X_P^2 &= 0.48 \quad \text{with D.F.} = 2 \quad \text{for litters of size 7} \\ X_P^2 &= 0.63 \quad \text{with D.F.} = 2 \quad \text{for litters of size 8} \\ X_P^2 &= 2.41 \quad \text{with D.F.} = 2 \quad \text{for litters of size 9} \\ X_P^2 &= 2.88 \quad \text{with D.F.} = 2 \quad \text{for litters of size 10} \\ X_P^2 &= 3.30 \quad \text{with D.F.} = 2 \quad \text{for litters of size 11} \; . \end{aligned} \tag{3.10}$$

The results in (3.9) clearly demonstrate that significant (α = .01) differences exist among the litter sizes for each treatment. However, they do no indicate the extent to which such differences are due to the fact that larger litters have more depletions because more animals are at risk.

Alternatively, none of the test statistics for treatment effects in (3.10) are significant (α = .10) although their respective magnitudes increase with litter size. A more refined approach for investigating treatment effects involves a bivariate combination of the information in the corresponding 2 x 3 tables in a manner analogous to that described by Cochran [1954] and Mantel [1963]. In particular, the methods described in Koch and Reinfurt [1973] were used for this purpose resulting in the combined Pearson χ^2-statistic X^2_{CP} = 6.59 with D.F. = 2 which suggests that significant (α = .05) differences exist between the effects of treatments A and B. However, this test statistic is primarily directed at the "average" differences between treatments across litter sizes, but provides no specific information regarding the treatment x litter size interaction which is suggested by the manner in which the statistics in (3.10) increase with litter size.

Thus, the results obtained from these simpler methods of analysis suggest that there are significant differences among litter sizes as well as between treatments. They do not, however, provide a clear indication of the nature of the litter size effect or its interaction with treatment. This same weakness is shared by other methods of analysis given for these data like that in GSK which are directed more at the overall number of depletions than a measure of the average rate of depletions. Thus, the primary advantage associated with the methods in Sections 3.1 and 3.2 is that they represent an effective approach for dealing with these types of difficult methodological questions which pertain to the analysis of clumped data.

ACKNOWLEDGMENTS

This research was supported in part by National Institutes of Health, Institute of General Medical Sciences through Grant GM-70004-04 and by the U.S. Bureau of the Census through Joint Statistical Agreement JSA 74-2. The authors would like to thank Rebecca Wesson for her conscientious typing of the manuscript.

REFERENCES

Berkson, J. [1968]. Application of minimum logit χ^2 estimate to a problem of Grizzle with a notation on the problem of no interaction. Biometrics 24, 75-95.

Bhapkar, V. P. and Koch, G. G. [1968a]. Hypotheses of 'no interaction' in multidimensional contingency tables. Technometrics 10, 107-123.

Bhapkar, V. P. and Koch, G. G. [1968b]. On the hypotheses of 'no interaction' in contingency tables. Biometrics 24, 567-594.

Cochran, W. G. [1954]. Some methods for strengthening the common χ^2 tests. Biometrics 10, 417-451.

Forthofer, R. N. and Koch, G. G. [1973]. An analysis for compounded functions of categorical data. Biometrics 29, 143-157.

Grizzle, J. E., Starmer, C. F. and Koch, G. G. [1969]. Analysis of categorical data by linear models. Biometrics 25, 489-504.

Kaplan, E. B. and Elston, R. C. [1973]. A subroutine package for maximum likelihood estimation (MAXLIK). North Carolina Institute of Statistics Mimeo Series, No. 823.

Kastenbaum, M. A. and Lamphiear, D. E. [1959]. Calculation of chi-square to calculate the no three-factor interaction hypothesis. Biometrics 15, 107-115.

Koch, G. G. and Reinfurt, D. W. [1973]. An analysis of the relationship between driver injury and vehicle age for automobiles involved in North Carolina accidents during 1966-1970. North Carolina Highway Safety Research Center Technical Report.

Koch, G. G. and Tolley, H. D. [1975]. A generalized modified-χ^2 analysis of categorical bacteria survival data from a complex dilution experiment. Biometrics 31.

Mantel, N. [1963]. Chi-square tests with one degree of freedom; extensions of the Mantel-Haenszel procedure. Journal of the American Statistical Association 58, 690-700.

Tolley, H. D. and Koch, G. G. [1974]. A two-stage approach to the analysis of longitudinal type categorical data. North Carolina Institute of Statistics Mimeo Series, No. 962.

Wald, A. [1943]. Tests of statistical hypotheses concerning several parameters when the number of observations is large. Transactions of the American Mathematical Society 54, 426-482.