JLTIPLE LINEAR REGRESSION VIEWPOINTS
VOLUME 16, NUMBER 1, SPRING 1988

A Perspective on Applications of Maximum Likelihood and Weighted Least Squares Procedures in the Context of Categorial Data Analysis

Andrew J. Bush
Baptist Memorial Hospital
Memphis, Tenn.

Pioneering technical contributions to the applied statistical literature by Grizzle, Starmer, and Koch (1969), Bishop (1969), Fienberg (1970), Goodman (1970), Koch and Reinfurt (1971), and, more recently, didactic contributions by Forthofer & Lehnen (1981) and by Kennedy (1983) have helped focus the attention of many research practitioners in the behavioral sciences on the potential for sophisticated analysis of categorical response data. In consequence, there is a growing awareness that a richer analysis can be performed on responses measured on the nominal or ordinal scale than is customarily permitted by simple crosstabulation and chi-square partitioning.

This awareness has led to the ever increasing popularity of strategies for the analysis of asymmetric, categorical data models--that is, models having at least one variable identified as a response variable. In particular, strategies that follow either the method of maximum likelihood (ML) in the Goodman tradition, such as log-linear (logit) and logistic regression analysis, or the method of weighted least squares in the Grizzle, Starmer, and Koch (GSK) tradition have been strongly gaining in acceptance.

Parenthetically, two points need now be made before proceeding to the main course of the narrative.

First, the strategies mentioned above also allow for the analysis of symmetric models--that is, models for which a dependent or response variable has not been identified. However, for the purpose of discussion, the focus here will be on asymmetric models.

Secondly, the GSK strategy subsumes an approach that is known by some as Minimum Chi-Square Estimation (cf. Aldrich and Nelson, 1984) and is a specific, direct, weighted least-squares approach employing categorical independent variables only. This point is made to call attention to the fact that the label,

weighted least-squares, is a general descriptor for any weighted regression procedure using any weighting factor whatsoever. Since differential selection of weighting schemes will produce different regression results, all weighted regression procedures are not equivalently effective. But, because of an unfortunate tendency to group any and all weighted procedures under a single label, the GSK procedure has had some undeserved bad press, in the form of guilt by association, from those who disparage the regression analysis of categorical data in general. The upshot of this digression is to admit that the GSK approach is a weighted regression approach with the further admission that it is fundamentally sound.

As might be expected, since the ML and the GSK approaches use different mathematical bases in their foundation, and thus can lead to differing statistical judgments, some dispute regarding their relative merits has begun to appear. Advocates of ML based strategies typically highly value log-linear and logistic regression analysis but look askance at the use of linear regression for the analysis of categorical outcomes. This position is particularly likely to develop amongst analysts who pursue log-linear problems from the mental framework of the Deming-Stephan iterative proportional fitting (IPF) algorithm (see Kennedy, 1983, for a particularly lucid description of

the algorithm).

By employing the IPF technique, a sound strategy in and of itself, it is unfortunately quite possible to miss the point that log-linear analysis is essentially a linear modeling process. More specifically, it is altogether too easy to overlook the tautology that log-linear models really are, in fact, linear models, and as such they can be structurally coded and resolved as linear models. Those familiar with the alternative to IPF, the Newton-Raphson iteratively reweighted regression algorithm for achieving ML estimates (see Haberman, 1978, for a full description), recognize the truth of this perspective much more readily.

In reality, that which separates ML from GSK analysis is not that one employs linear models and the other does not, nor is it that one employs a regression strategy and the other does not. Both, in fact, are rooted in a regression basis. What really separates the two is that their methods of implementing the regression strategy differ.

On the one hand, GSK seeks to achieve parameter estimates through minimizing a model's residual chi-square. It does so noniteratively under the mechanism of weighted least squares regression by adopting a weighting matrix formed as the inverse of the variance of a researcher specified response function (see Forthofer and Lehnen, 1981, for a very thorough

description).

ML, on the other hand, seeks to achieve parameter estimates by maximizing the likelihood function and does so iteratively under the mechanism of reweighted least squares regression. Per force, the weighting matrix, the basis matrix, and the form of the response variable for ML differ from those used under GSK.

Both strategies avoid the well-known problems that plague ordinary least squares in this context by not making untenable distributional assumptions. Neither assumes normality nor homogeneity of variance for the residual. Both assume independence and both typically assume a product-multinomial parent data distribution for asymmetric problems.

A Technical Overview Of The GSK And ML Categorical Data Analysis Strategies

To help fix the idea that both the GSK and ML procedures for analyzing categorical data are, in fact, regression based techniques, a summary overview of both procedures is offered on the following four pages. The technical description of each is highly condensed and is meant to give a reference point to the reader rather than a full, didactic exposition. The text underscores

that both procedures rest solidly on the foundation of weighted least squares (WLS). Pages six and seven describe major aspects of the GSK strategy while pages eight and nine deal with the ML approach.

The GSK Approach

Weighted Least Squares (WLS) analysis, employs a mathematical model that adopts the following notation:

1. $\underline{p}$ a vector of proportions. Each p_{ij} is computed as the ratio of a response frequency f_{ij} to $f_i = \sum_{j=1}^{J} f_{ij}$ where the subscript i indexes a particular independent variable level or combination of levels, the subscript j addresses a particular level of the response measure, and J denotes the number of levels present in the response measure. The elements of $\underline{p}$ are arranged so that the J proportions corresponding to a value of i are contiguous and in ascending order of j.
2. $\underline{A}$ a vector of contrast coefficients with elements a_j.
3. $\underline{Y}$ a vector of contrasts such that $Y = Ap$ for <u>additive</u> models. Each Y_i is formed as $Y_i = \sum a_j p_{ij}$. Alternatively intrinsically <u>multiplicative</u> models can be formulated by first taking the <u>natural log of</u> the p_{ij}. In this case, the vector $\underline{Y}$ is formed as $Y = A \ln(p)$. For such models, $Y_i = \sum a_j \ln(p_{ij})$.
4. $\underline{X}$ an independent variable coding matrix. For, WLS results to approximate those of a log-linear analysis, the matrix $\underline{X}$ is coded using effect codes (i.e., 1,0,-1).
5. $\underline{\beta}$ a vector of regression weights.
6. $\underline{\varepsilon}$ a vector of residuals.
7. $\underline{W}$ a matrix of weights such that $\underline{W} = V(Y)^{-1}$.

In the case of an <u>additive</u> model, $V(Y_i) = \frac{1}{f_i}\left[\sum_{j=1}^{r} a_j^2 p_{ij} - (\sum_{j=1}^{r} a_j p_{ij})^2\right]$

Should r=2 and A = [1 0] or A = [0 1], $V(Y_i) = \frac{p_{ij}(1-p_{ij})}{f_i}$ for j=1 or j=2.

In the case of a <u>multiplicative</u> model, $V(Y_i) = \sum_{j=1}^{r} \frac{a_j^2}{f_{ij}} - \frac{1}{f_i}(\sum_{j=1}^{r} a_j)^2$. Here, should r=2 and A = [1 -1] or A = [-1 1], (the logit function), then it follows that $V(Y_i) = \frac{1}{f_i p_{ij}(1-p_{ij})}$ for either j=1 or j=2.

Using these conventions, the regression model can be written as:

$$Y = X\beta + \epsilon$$
$$b = (X^T W X)^{-1} (X^T W Y)$$
$$V(b) = (X^T W X)^{-1}$$
$$\hat{Y} = Xb$$
$$V(\hat{Y}) = X (X^T W X)^{-1} X^T$$

The <u>residual</u> chi-square for such models is:

$$\chi^2 = (Y - Xb)^T W (Y - Xb)$$

with df = k - m

where k = the number of independent cells (i.e., rows in X)

and m = the number of parameters (i.e., columns in X)

Given a contrast matrix C that has dimensions c x m, component chi-squares (i.e., corresponding to the general linear <u>hypothesis</u> $C\beta = 0$) can be computed as:

$$\chi^2 = (Cb)^T [C(X^T W X)^{-1} C^T]^{-1} Cb$$

with df = c

Approximations to component chi-squares, can also be computed by taking the difference in residual chi-squares for competing models with <u>df</u> equal to the difference in the respective number of parameters. This approximation method is not as effective here as it is in log-linear analyses since the chi-square estimates are the classical Pearsonian rather than the maximum likelihood ratio chi-squares developed by Fisher and are, consequently, not precisely additive.

Iterative Weighted Least Squares (WLS) can be used to achieve Maximum Likelihood (ML) estimates. The strategy assumes the following notational conventions:

1. A diagonal matrix $\underline{F}$ of dimensionality (kr x kr) where k is the number of independent variable cells and r is the number of response variable levels. The elements of $\underline{F}$ are individual f_i where $1 <= i <= kr$. They are arranged on the major diagonal so that the order of rotaton is through the response levels for a particular independent variable cell before the next cell is represented.

2. A diagonal matrix $\underline{E}$ whose entries e_i are the expected frequencies for a given model in correspondence to the f_i.

3. A design matrix $\underline{X}$ of dimensionality (kr x m) where m is equal to the sum $(k-1)+(r-1)+(k-1)*(r-1)$. Note that m represents the total component degrees of freedom in a given model excluding the intercept (or grand mean) which is not coded. The design matrix $\underline{X}$ is composed of effect codes (1,0,-1) and is formed as:

   a. The first $k-1$ columns of $\underline{X}$ are effect codes on the independent variables-- each row of which is replicated contiguously r times.

   b. The next $r-1$ columns of $\underline{X}$ are formed by block replicating effect codes on the response measure k times. Each block is of dimensionality r x $r-1$.

   c. The remaining $(k-1)*(r-1)$ columns represent the independent-dependent variable interaction terms and are formed by multiplication of the corresponding prior columns.

4. The subscript c represents the current iteration and the subscript p represents the prior iteration.

5. Vector $\underline{Y}$ = diag $[\ln(E_p) + (F - E_p)E_p^{-1}]$. On the first iteration, this procedure is replaced by computing each element $\underline{Y}$ to be $Y_i = \ln(e_i)$ where $e_i = f_i + .05$.

6. A matrix $\underline{D}$ of the same dimensionality as $\underline{X}$ formed by kr row replicates of the vector $\underline{d}$ with elements d_j $(1 <= j <= m)$

where
$$d_j = \frac{\sum_{i=1}^{kr} X_{ij}\, e_i}{\sum_{i=1}^{kr} e_i}$$

given the e_i are from the prior iteration.

The iterative process, given $\underline{X}$ and $\underline{E}$, is as follows:

1. Compute $\underline{Y}_c$ as described
2. Compute $\underline{D}_c$ as described
3. Generate the matrix $\underline{A} = \underline{X} - \underline{D}_c$
4. Estimate the regression weights β as

$$b_c = [A_c^T E_p A_c]^{-1} A_c^T E_p Y_c$$

$$a_c = \ln \left[\frac{\sum_{i=1}^{kr} f_i}{\sum_{i=1}^{kr} e^{\sum_{j=1}^{m} b_j x_{ij}}} \right]$$

5. Estimate the i^{th} element of E_c as $e^{a_c + \sum_{j=1}^{m} b_j x_{ij}}$
6. If the estimates b_c converge on b_p then stop iteration otherwise return to step 1.

Given convergence, the following additional estimates can be made:

1. $V(b) = [A_c^T E_c A_c]^{-1}$
2. Standardized residuals are $(f_i - e_i)/\sqrt{e_i}$
3. Residual $L^2 = 2 \sum_{i=1}^{kr} f_i \ln\left(\frac{f_i}{e_i} \right)$
4. Residual $\chi^2 = \sum_{i=1}^{kr} \frac{(f_i - e_i)^2}{e_i}$

both with $\underline{df} = kr - m - 1$

As the reader can readily see, both approaches permit point and interval estimation of regression parameters. To help profile how the strategies compare with one another, their relative merit from the author's point of view will now be examined along several dimensions. Those dimensions are:

1) Ability to deal with symmetric models.
2) Facility for testing hypotheses.
3) Statistical properties of estimators.
4) Relative computational requirements.
5) Ease of interpretation of estimators.
6) Robustness with respect to extreme values.
7) Capacity for handling interval variables.

Symmetric Models. With regard to doing data analysis where no individual variable is perceived to be a response (dependent) variable, the ML method has a clear edge. In fact, log-linear analysis, having its roots in the field of sociological methodology, a field that does not often enjoy the luxury of experimental manipulation of independent variables, is exceptionally well-geared for coping with marginal and partial associations among variables.

In contrast, the GSK approach, an approach that emanates from the biostatistical world, is focused directly on exploring the effects of one or more independent on one or more dependent variables. Unlike the log-linear strategy, GSK forces selection of a response variable. This does not mean that the GSK approach can not handle symmetric problems--it can. However, an analyst must systematically rotate through a problem's variables choosing different variables, individually, as the response measure. Consequently, the GSK method is not as desirable in such a context.

Facility for Testing Hypotheses. Assuming the asymmetric environment for the remainder of this narrative, how do the strategies compare on the basis of testing hypotheses? In this writer's opinion, the GSK approach is probably stronger but not overpoweringly so. GSK, on the surface, appears to have far greater flexibility because the analyst is permitted to establish nearly any linear combination on nearly any transformation of the response measure. Such flexibility permits definition of a response function in terms of raw proportions, or logged proportions (the latter leading directly to odds ratios), or even exponentiated proportions.

In comparison, the log-linear approach forces a definition of the response function in terms of logged proportions. However, what is often overlooked during a log-linear analysis is that expected frequencies are generated and that the analyst is free to establish any desired transformation and linear combination on those frequencies he or she wishes. This implies that the ML method can be as rich analytically as the GSK method (cf. Haber, 1984). In fairness, though, the more extended mode of analysis under ML is not typical and is more mechanically difficult.

Statistical Properties. With respect to the statistical properties of the estimators produced by GSK and ML, a slight edge has to be awarded ML since the ML estimators are well-known to be asymptotically consistent and relatively efficient. What is not as well known is that the GSK estimators are similarly asymptotically consistent and, for that matter, asymptotically equivalent to ML estimators. They are, in fact, best asymptotic normal estimators (BAN).

For fully saturated models of any sample size, the two methods deliver identical results. For unsaturated models on large samples, differences in the estimators tend to be trivial. However, as sample sizes decrease,

the GSK and the ML estimators can be disparate with the ML estimators tending to have smaller variance. The question of how large is large enough to feel fairly comfortable that similar results will be afforded by both strategies is not precisely known. However, it is generally recommended that samples be of sufficient size before employing either approach. For specific guidelines under GSK, the reader is referred to Forthofer and Lehnen (1981) and, for guidelines under ML, to Haberman (1978).

Computational Requirements. From a computational perspective, GSK has a clear edge. In the first place, it is non-iterative. In the second, its basis matrix is a factor of r*r smaller where r denotes the number of categories present in the response variable. For problems involving polytomous response measures, computational resource requirements heavily favor GSK. While such considerations may not be critical for mainframe applications, the resource implications for microcomputing are clear.

Ease of Interpretation. With regard to estimator interpretability, ML estimates are slightly easier for a novice to make sense of if a canned log-linear strategy is being employed. This is the case because

the parameters are conceptually well identified in the paradigm of analysis of variance effects on logged expected cell frequencies. If, however, the more flexible regression coding scheme afforded by the Newton-Raphson strategy is employed to deviate from traditional effect definitions, this edge evaporates and both ML and GSK estimates must be carefully identified by the analyst.

Robustness for Extreme Values. From the perspective of extreme values, the GSK and the ML strategies share common problems. Both must cope with empty cells by either making a numeric replacement or collapsing categories. Further, both rely on having large samples to effect robustness in the statistical properties of their estimators. From this author's viewpoint, neither procedure has an edge with regard to this problem. However, it should be noted that it is recommended that the GSK approach engage a log transformation on proportions when proportions are extreme rather than operating upon them in their native metric (see Forthofer & Lehnen, 1981). Intuitively, the same caveat should apply to followup contrasts on ML estimates.

Interval Independent Variables. With regard to interval independent variables, one variant of ML, namely logistic regression analysis, has a distinctive advantage. It has the capacity for coping with a mix of both categorical and continuous variables with the provision that the response measure be a dichotomous variable.

Neither the GSK nor traditional log-linear ML methods can duplicate this capacity. Even so, an analyst could approach the situation of interval variables with either log-linear or GSK analysis by meaningfully categorizing all interval variables present.

Synthesis. Given this profile, which procedure then is preferable? From the author's perspective neither completely dominates the other. Both are powerful and are well worth mastering.

Should the research purpose be to examine marginal and partial associations symmetrically, the ML approach embodied by log-linear analysis is preferable. Should the research purpose be to test hypotheses on response level proportions or on complex functions, the GSK approach is preferable. If interval level independent

variables are present and recoding is not desirable, the logistic regression ML approach is promising--providing no more than two levels are present in the response variable.

Should computing facilities be highly restricted, the GSK approach can be preferable. If the analyst is unsophisticated with respect to the analysis of linear models, a traditional log-linear analysis will be easier to pursue. If sample sizes are small or empty cells are present, neither strategy is particularly safe. If extreme proportions are present, both approaches should make appropriate adjustments.

In the final analysis, both approaches have specific strengths as well as detractions. Both offer strong analytic capabilities and both belong in our repertoire.

An Analysis of Hypothetical Data By ML And By GSK

For the purpose of illustrating the similarity of the two methods in an applied scenario and for the purpose of demonstrating their versatility, the following simple numeric example is offered. The data shown below were constructed by John J. Kennedy, of The

Ohio State University, as a didactic example to show how effect contrasts might be estimated through chi-square partitioning. With his kind permission, the data will be employed here to show (1) how both ML and GSK can be used to estimate linear and quadratic effects and (2) how both the ML and GSK procedures can pursue traditional log-linear effects.

The data are given in Table 1 and consist of frequency counts that have been crosstabulated on the basis of student sex (A_1= Males, A_2= Females), an unspecified treatment variable (B_1= Treatment, B_2= Control), and a trichotomous outcome measure (C_1= Poor, C_2= Satisfactory, and C_3= Good).

Table 1. A Hypothetical 2x2x3 Data Example.

		Outcome Description:			
		Poor	Satisfactory	Good	Sum
Sex	Treatment				
M	T	5	19	4	28
M	C	3	6	13	22
F	T	6	16	6	28
F	C	2	8	12	22
Sum		16	49	35	100

Page 19 demonstrates a linear and quadratic effect coding setup used as input to an author prepared Newton-Raphson ML program that has been designed to teach the flow of the ML procedure. The input consists of (1) the number of rows in the regression basis matrix, (2) the number of columns in that matrix--note the omission of a unit vector for the grand mean, (3) the basis matrix, itself, arranged in column order:

a) Sex vector.

b) Treatment vector.

c) Sex x Treatment.

d) Linear Response Contrast.

e) Quadratic Response Contrast.

f) Linear Effect of Sex.

g) Quadratic Effect of Sex.

h) Linear Effect of Treatment.

i) Quadratic Effect of Treatment.

j) Linear Effect of Sex x Treatment.

k) Quadratic Effect of Sex x Treatment.

and (4) the raw frequencies themselves with the response variable rotating most rapidly, followed by treatment, and sex in that order. Pages 20 and 21 show the ML analysis with page 21 being the more interesting since it delivers parameter estimates. Pages 22 and 23 show the corresponding GSK analysis with page 22 delivering the linear analysis and 23, the quadratic.

ML Analysis of 2x2x3 Data Set Using Linear & Quadratic Codings

```
12
11
 1   1   1 -.5 -.333333    -.5 -.333333   -.5 -.333333  -.5 -.333333
 1   1   1   0  .666667      0  .666667     0  .666667    0  .666667
 1   1   1  .5 -.333333     .5 -.333333    .5 -.333333   .5 -.333333
 1  -1  -1 -.5 -.333333    -.5 -.333333    .5  .333333   .5  .333333
 1  -1  -1   0  .666667      0  .666667     0 -.666667    0 -.666667
 1  -1  -1  .5 -.333333     .5 -.333333   -.5  .333333  -.5  .333333
-1   1  -1 -.5 -.333333     .5  .333333   -.5 -.333333   .5  .333333
-1   1  -1   0  .666667      0 -.666667     0  .666667    0 -.666667
-1   1  -1  .5 -.333333    -.5  .333333    .5 -.333333  -.5  .333333
-1  -1   1 -.5 -.333333     .5  .333333    .5  .333333  -.5 -.333333
-1  -1   1   0  .666667      0 -.666667     0 -.666667    0  .666667
-1  -1   1  .5 -.333333    -.5  .333333   -.5  .333333   .5 -.333333
5
19
4
3
6
13
6
16
6
2
8
12
```

ML Analysis of 2x2x3 Data Set Using Linear & Quadratic Codings

Cell Frequencies Iteration is 4

cell =>	1	obs freq =>	5.0000	exp freq =>	5.0000
cell =>	2	obs freq =>	19.0000	exp freq =>	19.0000
cell =>	3	obs freq =>	4.0000	exp freq =>	4.0000
cell =>	4	obs freq =>	3.0000	exp freq =>	3.0000
cell =>	5	obs freq =>	6.0000	exp freq =>	6.0000
cell =>	6	obs freq =>	13.0000	exp freq =>	13.0000
cell =>	7	obs freq =>	6.0000	exp freq =>	6.0000
cell =>	8	obs freq =>	16.0000	exp freq =>	16.0000
cell =>	9	obs freq =>	6.0000	exp freq =>	6.0000
cell =>	10	obs freq =>	2.0000	exp freq =>	2.0000
cell =>	11	obs freq =>	8.0000	exp freq =>	8.0000
cell =>	12	obs freq =>	12.0000	exp freq =>	12.0000

Dvector Iteration is 4

column =>	1	value =>	-0.0000
column =>	2	value =>	0.1200
column =>	3	value =>	0.0000
column =>	4	value =>	0.0950
column =>	5	value =>	0.1567
column =>	6	value =>	-0.0050
column =>	7	value =>	0.0100
column =>	8	value =>	-0.1050
column =>	9	value =>	0.1700
column =>	10	value =>	-0.0050
column =>	11	value =>	0.0500

Amatrix Iteration is 4

1.000	0.880	1.000	-0.595	-0.490	-0.495	-0.343	-0.395	-0.503	-0.495	-0
1.000	0.880	1.000	-0.095	0.510	0.005	0.657	0.105	0.497	0.005	0
1.000	0.880	1.000	0.405	-0.490	0.505	-0.343	0.605	-0.503	0.505	-0
1.000	-1.120	-1.000	-0.595	-0.490	-0.495	-0.343	0.605	0.163	0.505	0
1.000	-1.120	-1.000	-0.095	0.510	0.005	0.657	0.105	-0.837	0.005	-0
1.000	-1.120	-1.000	0.405	-0.490	0.505	-0.343	-0.395	0.163	-0.495	0
-1.000	0.880	-1.000	-0.595	-0.490	0.505	0.323	-0.395	-0.503	0.505	0
-1.000	0.880	-1.000	-0.095	0.510	0.005	-0.677	0.105	0.497	0.005	-0
-1.000	0.880	-1.000	0.405	-0.490	-0.495	0.323	0.605	-0.503	-0.495	0
-1.000	-1.120	1.000	-0.595	-0.490	0.505	0.323	0.605	0.163	-0.495	-0
-1.000	-1.120	1.000	-0.095	0.510	0.005	-0.677	0.105	-0.837	0.005	0
-1.000	-1.120	1.000	0.405	-0.490	-0.495	0.323	-0.395	0.163	0.505	-0

```
Analysis of 2x2x3 Data Set Using Linear & Quadratic Codings

s Iteration is  4

ercept is     1.917424      old value was     1.917424

umn =>  1   A      value =>    -0.018176     Change     0.000000
umn =>  2   B      value =>     0.131955     Change     0.000000
umn =>  3   AB     value =>    -0.051147     Change    -0.000000
umn =>  4   C1     value =>     0.758738     Change     0.000000
umn =>  5   C2     value =>     0.719449     Change     0.000000
umn =>  6   AC1    value =>    -0.137141     Change    -0.000000
umn =>  7   AC2    value =>    -0.016173     Change    -0.000000
umn =>  8   BC1    value =>    -0.870310     Change    -0.000000
umn =>  9   BC2    value =>     0.494252     Change    -0.000000
umn => 10   ABC1   value =>     0.025570     Change     0.000000
umn => 11   ABC2   value =>     0.249045     Change     0.000000

 of changes     0.000000

iance Iteration is  4

0152 0.0017-0.0027 0.0044 0.0013-0.0130-0.0100-0.0023-0.0041 0.0151-0.0015
0017 0.0152-0.0002 0.0151-0.0015-0.0023-0.0041-0.0130-0.0100 0.0044 0.0013
0027-0.0002 0.0152-0.0023-0.0041 0.0151-0.0015 0.0044 0.0013-0.0130-0.0100
0044 0.0151-0.0023 0.1111 0.0195-0.0035-0.0066-0.0131-0.0226 0.0181 0.0034
0013-0.0015-0.0041 0.0195 0.0532-0.0066 0.0011-0.0226-0.0143 0.0034 0.0013
0130-0.0023 0.0151-0.0035-0.0066 0.1111 0.0195 0.0181 0.0034-0.0131-0.0226
0100-0.0041-0.0015-0.0066 0.0011 0.0195 0.0532 0.0034 0.0013-0.0226-0.0143
0023-0.0130 0.0044-0.0131-0.0226 0.0181 0.0034 0.1111 0.0195-0.0035-0.0066
0041-0.0100 0.0013-0.0226-0.0143 0.0034 0.0013 0.0195 0.0532-0.0066 0.0011
0151 0.0044-0.0130 0.0181 0.0034-0.0131-0.0226-0.0035-0.0066 0.1111 0.0195
0015 0.0013-0.0100 0.0034 0.0013-0.0226-0.0143-0.0066 0.0011 0.0195 0.0532


teration is  4

umn =>  1     value =>     3.2242
umn =>  2     value =>    12.4921
umn =>  3     value =>     4.1241
umn =>  4     value =>     6.0796
umn =>  5     value =>    14.6828
umn =>  6     value =>     0.7101
umn =>  7     value =>     2.3169
umn =>  8     value =>    -6.2753
umn =>  9     value =>    20.3232
umn => 10     value =>    -0.9058
umn => 11     value =>     4.5616

rsonian       0.0000
herian        0.0000
```

The Pattern Matrix X as Entered

1.00	1.00	1.00	1.00
1.00	1.00	-1.00	-1.00
1.00	-1.00	1.00	-1.00
1.00	-1.00	-1.00	1.00

The Parameter Coefficient Matrix:

0.25	0.25	0.25	0.25
0.25	0.25	-0.25	-0.25
0.25	-0.25	0.25	-0.25
0.25	-0.25	-0.25	0.25

The Frequencies as Entered

CATEGORY:		
1	2	3
5	19	4
3	6	13
6	16	6
2	8	12

CONTRAST: -1.00 0.00 1.00

PARAMETER	LOG EST	LOG SE	ODDS EST	ODDS SE	Z ESTIMATE
INTERCEPT	0.769	0.333	2.136	1.396	2.277
AC1	-0.137	0.333	0.872	1.396	-0.412
BC1	-0.870	0.333	0.419	1.396	-2.612
ABC1	0.026	0.333	1.026	1.396	0.077

PERFECT FIT --- SATURATED MODEL

RESIDUAL CHI-SQUARE = 0.000 DF = 0 ALPHA = 1.00

LOG-P FUNCTION	PREDICTED	RESIDUAL
-0.223	-0.223	0.000
1.466	1.466	0.000
0.000	0.000	0.000
1.792	1.792	0.000

The Pattern Matrix X as Entered

.00	1.00	1.00	1.00
.00	1.00	-1.00	-1.00
.00	-1.00	1.00	-1.00
.00	-1.00	-1.00	1.00

The Parameter Coefficient Matrix:

.25	0.25	0.25	0.25
.25	0.25	-0.25	-0.25
.25	-0.25	0.25	-0.25
25	-0.25	-0.25	0.25

The Frequencies as Entered

CATEGORY:

1	2	3
5	19	4
3	6	13
6	16	6
2	8	12

RAST: -0.50 1.00 -0.50

METER RCEPT	LOG EST	LOG SE	ODDS EST	ODDS SE	Z ESTIMATE
	0.719	0.231	2.053	1.259	3.120
	-0.016	0.231	0.984	1.259	-0.070
	0.494	0.231	1.639	1.259	2.143
	0.249	0.231	1.283	1.259	1.080

PERFECT FIT --- SATURATED MODEL

RESIDUAL CHI-SQUARE = 0.000 DF = 0 ALPHA = 1.00

LOG-P FUNCTION	PREDICTED	RESIDUAL
1.447	1.447	0.000
-0.040	-0.040	0.000
0.981	0.981	0.000
0.490	0.490	0.000

Collecting the effect estimates from the runs just presented lets us produce Table 2. Note that two separate analyses had to be performed by GSK to produce first the linear and then the quadratic results.

Table 2. Summary of ML & GSK Analysis of Linear & Quadratic Effects in the 2x2x3 Example.

	ML			GSK		
Effect	b	SE	Page	b	SE	Page
AC_1	-.14	.33	21	-.14	.33	22
AC_2	-.02	.23	21	-.02	.23	23
BC_1**	-.87	.33	21	-.87	.33	22
BC_2*	.49	.23	21	.49	.23	23
ABC_1	.03	.33	21	.03	.33	22
ABC_2	.25	.23	21	.25	.23	23

** p .01

* p .05

Clearly the two sets of results are isomorphic with each revealing both a linear and quadratic effect for the treatment variable on the response frequencies. With respect to the linear trend, the odds favoring a response of "good" over a response of "poor" are better in the control group than in the treatment.

With respect to the quadratic trend, the treatment group average odds favoring a "satisfactory" response over the other two response categories are better than the corresponding odds for the control condition. Obviously, if this were a true research situation, an analyst would suddenly get gray hair but the data do serve the purpose of illustration.

Repeating the exercise with linear codings established to produce traditional log-linear parameters, the ML input file is shown on page 26 and follows exactly the same pattern as before. This time, however, the linear and quadratic codes give way to average effect codes.

Pages 27 and 28 reproduce the results from the ML analysis with page 28 being the more interesting. The GSK output is shown on pages 29, 30, and 31. This time three runs were made under GSK in order to directly estimate the parameters associated with the third level of the response variable. These could, admittedly, have been determined by subtraction. However, the variance estimates for the parameters on page 31 would have had to have been inferred rather than obtained from inspection.

```
ML Analysis of 2x2x3 Data Set Using Log-Linear Codings

12
11
 1   1   1   1   0    1   0    1   0    1   0
 1   1   1   0   1    0   1    0   1    0   1
 1   1   1  -1  -1   -1  -1   -1  -1   -1  -1
 1  -1  -1   1   0    1   0   -1   0   -1   0
 1  -1  -1   0   1    0   1    0  -1    0  -1
 1  -1  -1  -1  -1   -1  -1    1   1    1   1
-1   1  -1   1   0   -1   0    1   0   -1   0
-1   1  -1   0   1    0  -1    0   1    0  -1
-1   1  -1  -1  -1    1   1   -1  -1    1   1
-1  -1   1   1   0   -1   0   -1   0    1   0
-1  -1   1   0   1    0  -1    0  -1    0   1
-1  -1   1  -1  -1    1   1    1   1   -1  -1
5
19
4
3
6
13
6
16
6
2
8
12
12
11
```

nalysis of 2x2x3 Data Set Using Log-Linear Codings

Cell Frequencies Iteration is 4

. =>	1	obs freq =>	5.0000	exp freq =>	5.0000
. =>	2	obs freq =>	19.0000	exp freq =>	19.0000
. =>	3	obs freq =>	4.0000	exp freq =>	4.0000
. =>	4	obs freq =>	3.0000	exp freq =>	3.0000
. =>	5	obs freq =>	6.0000	exp freq =>	6.0000
. =>	6	obs freq =>	13.0000	exp freq =>	13.0000
. =>	7	obs freq =>	6.0000	exp freq =>	6.0000
. =>	8	obs freq =>	16.0000	exp freq =>	16.0000
. =>	9	obs freq =>	6.0000	exp freq =>	6.0000
. =>	10	obs freq =>	2.0000	exp freq =>	2.0000
. =>	11	obs freq =>	8.0000	exp freq =>	8.0000
. =>	12	obs freq =>	12.0000	exp freq =>	12.0000

tor Iteration is 4

mn =>	1	value =>	-0.0000
mn =>	2	value =>	0.1200
mn =>	3	value =>	0.0000
mn =>	4	value =>	-0.1900
mn =>	5	value =>	0.1400
mn =>	6	value =>	0.0100
mn =>	7	value =>	0.0200
mn =>	8	value =>	0.2100
mn =>	9	value =>	0.3600
mn =>	10	value =>	0.0100
mn =>	11	value =>	0.0800

trix Iteration is 4

.000	0.880	1.000	1.190	-0.140	0.990	-0.020	0.790	-0.360	0.990	-0.080
.000	0.880	1.000	0.190	0.860	-0.010	0.980	-0.210	0.640	-0.010	0.920
.000	0.880	1.000	-0.810	-1.140	-1.010	-1.020	-1.210	-1.360	-1.010	-1.080
.000	-1.120	-1.000	1.190	-0.140	0.990	-0.020	-1.210	-0.360	-1.010	-0.080
.000	-1.120	-1.000	0.190	0.860	-0.010	0.980	-0.210	-1.360	-0.010	-1.080
.000	-1.120	-1.000	-0.810	-1.140	-1.010	-1.020	0.790	0.640	0.990	0.920
.000	0.880	-1.000	1.190	-0.140	-1.010	-0.020	0.790	-0.360	-1.010	-0.080
.000	0.880	-1.000	0.190	0.860	-0.010	-1.020	-0.210	0.640	-0.010	-1.080
.000	0.880	-1.000	-0.810	-1.140	0.990	0.980	-1.210	-1.360	0.990	0.920
.000	-1.120	1.000	1.190	-0.140	-1.010	-0.020	-1.210	-0.360	0.990	-0.080
.000	-1.120	1.000	0.190	0.860	-0.010	-1.020	-0.210	-1.360	-0.010	0.920
.000	-1.120	1.000	-0.810	-1.140	0.990	0.980	0.790	0.640	-1.010	-1.080

```
ML Analysis of 2x2x3 Data Set Using Log-Linear Codings

Bwts Iteration is  4

intercept is     1.917425      old value was     1.917425

column =>  1  A      value =>    -0.018176     Change      0.000000
column =>  2  B      value =>     0.131955     Change      0.000000
column =>  3  AB     value =>    -0.051147     Change     -0.000000
column =>  4  C1     value =>    -0.619185     Change     -0.000000
column =>  5  C2     value =>     0.479633     Change      0.000000
column =>  6  AC1    value =>     0.073962     Change      0.000000
column =>  7  AC2    value =>    -0.010782     Change     -0.000000
column =>  8  BC1    value =>     0.270404     Change      0.000000
column =>  9  BC2    value =>     0.329501     Change     -0.000000
column => 10  ABC1   value =>    -0.095800     Change     -0.000000
column => 11  ABC2   value =>     0.166030     Change      0.000000

Sum of changes     0.000000

Variance Iteration is  4

 0.0152 0.0017-0.0027-0.0026 0.0008 0.0098-0.0067 0.0025-0.0027-0.0070-0.
 0.0017 0.0152-0.0002-0.0070-0.0010 0.0025-0.0027 0.0098-0.0067-0.0026 0.
-0.0027-0.0002 0.0152 0.0025-0.0027-0.0070-0.0010-0.0026 0.0008 0.0098-0.
-0.0026-0.0070 0.0025 0.0402-0.0183-0.0029 0.0019-0.0124 0.0107 0.0058-0.
 0.0008-0.0010-0.0027-0.0183 0.0236 0.0019 0.0005 0.0107-0.0064-0.0014 0.
 0.0098 0.0025-0.0070-0.0029 0.0019 0.0402-0.0183 0.0058-0.0014-0.0124 0.
-0.0067-0.0027-0.0010 0.0019 0.0005-0.0183 0.0236-0.0014 0.0006 0.0107-0.
 0.0025 0.0098-0.0026-0.0124 0.0107 0.0058-0.0014 0.0402-0.0183-0.0029 0.
-0.0027-0.0067 0.0008 0.0107-0.0064-0.0014 0.0006-0.0183 0.0236 0.0019 0.
-0.0070-0.0026 0.0098 0.0058-0.0014-0.0124 0.0107-0.0029 0.0019 0.0402-0.
-0.0010 0.0008-0.0067-0.0014 0.0006 0.0107-0.0064 0.0019 0.0005-0.0183 0.

XY Iteration is  4

column =>  1     value =>     3.2242
column =>  2     value =>    12.4921
column =>  3     value =>     4.1241
column =>  4     value =>   -12.1592
column =>  5     value =>    15.9446
column =>  6     value =>    -1.4202
column =>  7     value =>     2.7652
column =>  8     value =>    12.5507
column =>  9     value =>    36.7601
column => 10     value =>     1.8116
column => 11     value =>     7.7483

Pearsonian      0.0000
Fisherian       0.0000
```

GSK Log-linear: C1 odds p 29

The Pattern Matrix X as Entered

)0	1.00	1.00	1.00
)0	1.00	-1.00	-1.00
)0	-1.00	1.00	-1.00
)0	-1.00	-1.00	1.00

The Parameter Coefficient Matrix:

25	0.25	0.25	0.25
25	0.25	-0.25	-0.25
25	-0.25	0.25	-0.25
25	-0.25	-0.25	0.25

The Frequencies as Entered

CATEGORY:

1	2	3
5	19	4
3	6	13
6	16	6
2	8	12

RAST: 0.67 -0.33 -0.33

METER	LOG EST	LOG SE	ODDS EST	ODDS SE	Z ESTIMATE
RCEPT	-0.619	0.200	0.538	1.222	-3.090
	0.074	0.200	1.077	1.222	0.369
	0.270	0.200	1.310	1.222	1.349
	-0.096	0.200	0.909	1.222	-0.478

PERFECT FIT --- SATURATED MODEL

RESIDUAL CHI-SQUARE = 0.000 DF = 0 ALPHA = 1.00

LOG-P FUNCTION	PREDICTED	RESIDUAL
-0.371	-0.371	0.000
-0.720	-0.720	0.000
-0.327	-0.327	0.000
-1.059	-1.059	0.000

GSK Log-linear: C2 odds p 30

The Pattern Matrix X as Entered

1.00	1.00	1.00	1.00
1.00	1.00	-1.00	-1.00
1.00	-1.00	1.00	-1.00
1.00	-1.00	-1.00	1.00

The Parameter Coefficient Matrix:

0.25	0.25	0.25	0.25
0.25	0.25	-0.25	-0.25
0.25	-0.25	0.25	-0.25
0.25	-0.25	-0.25	0.25

The Frequencies as Entered

CATEGORY:		
1	2	3
5	19	4
3	6	13
6	16	6
2	8	12

CONTRAST: -0.33 0.67 -0.33

PARAMETER	LOG EST	LOG SE	ODDS EST	ODDS SE	Z ESTIMATE
INTERCEPT	0.480	0.154	1.615	1.166	3.120
AC2	-0.011	0.154	0.989	1.166	-0.070
BC2	0.330	0.154	1.390	1.166	2.143
ABC2	0.166	0.154	1.181	1.166	1.080

PERFECT FIT --- SATURATED MODEL

RESIDUAL CHI-SQUARE = 0.000 DF = 0 ALPHA = 1.00

LOG-P FUNCTION	PREDICTED	RESIDUAL
0.964	0.964	0.000
-0.027	-0.027	0.000
0.654	0.654	0.000
0.327	0.327	0.000

Again collecting the computed results produces Table 3. Once more the profile is consistent.

Table 3. Summary of ML & GSK Analysis of Log-linear Effects in the 2x2x3 Example.

	ML			GSK		
Effect	b	SE	Page	b	SE	Page
AC_1	.07	.20	28	.07	.20	29
AC_2	-.01	.15	28	-.01	.15	30
AC_3	-.06	.16		-.06	.16	31
BC_1	.27	.20	28	.27	.20	29
BC_2*	.33	.15	28	.33	.15	30
BC_3**	-.60	.16		-.60	.16	31
ABC_1	-.10	.20	28	-.10	.20	29
ABC_2	.17	.15	28	.17	.15	30
ABC_3	-.07	.16		-.07	.16	31

** $p < .01$

* $p < .05$

Once more we clearly have identical results but now in terms of log-linear estimates. By way of interpretation, the significant BC_2 term indicates that the geometric average odds favoring a "satisfactory" response over all possible response categories are

GSK Log-linear: C3 odds p 31

The Pattern Matrix X as Entered

1.00	1.00	1.00	1.00
1.00	1.00	-1.00	-1.00
1.00	-1.00	1.00	-1.00
1.00	-1.00	-1.00	1.00

The Parameter Coefficient Matrix:

0.25	0.25	0.25	0.25
0.25	0.25	-0.25	-0.25
0.25	-0.25	0.25	-0.25
0.25	-0.25	-0.25	0.25

The Frequencies as Entered

CATEGORY:

1	2	3
5	19	4
3	6	13
6	16	6
2	8	12

CONTRAST: -0.33 -0.33 0.67

PARAMETER	LOG EST	LOG SE	ODDS EST	ODDS SE	Z ESTIMATE
INTERCEPT	0.140	0.165	1.150	1.179	0.846
AC3	-0.063	0.165	0.939	1.179	-0.383
BC3	-0.600	0.165	0.549	1.179	-3.639
ABC3	-0.070	0.165	0.932	1.179	-0.426

PERFECT FIT --- SATURATED MODEL

RESIDUAL CHI-SQUARE = 0.000 DF = 0 ALPHA = 1.00

LOG-P FUNCTION	PREDICTED	RESIDUAL
-0.594	-0.594	0.000
0.747	0.747	0.000
-0.327	-0.327	0.000
0.732	0.732	0.000

stronger for the treatment group than the controls. The significant BC_3 term indicates that the average odds favoring a "good" response are better in the control condition. The results are consistent with the findings from the linear-quadratic analysis but reveal a slightly different aspect of the data based on the differential coding. Again, thankfully, the results are fictitious.

Concluding Remarks

The author hopes that a relatively convincing case has been built for embracing both the ML and GSK technologies and for appreciating that both are fundamentally regression based strategies. Further, he hopes that the point has been adequately made that to argue which is better is, at best, a contextually bound issue which begs the question for a universal answer.

Certainly, much more could have been discussed regarding relative applications, for example, with respect to nested and blocking design or with respect to followups to omnibus tests. These matters are relevant and important but beyond the scope of the material presented here. Obviously the application arena is large and the application tools are superb.

References

1. Aldrich, J.H., & Nelson, F.D., *Linear probability, logit, and probit models*. Beverly Hills: Sage University Press, 1984.

2. Bishop, Y.M.M., Full contingency tables, logits, and split contingency tables. *Biometrics*, 1969, *25*, 383-400.

3. Fienberg, S.E., The analysis of multimensional contingency tables. *Ecology*, 1970, *51*, 419-433.

4. Forthofer, R.N., & Lehnen, R.G., *Public program analysis: A new categorical data approach*. Belmont: Lifetime Learning Publications, 1981.

5. Goodman, L.A., The multivariate analysis of qualitative data: Interactions among multiple classifications. *Journal of the American Statistical Association*, 1970, *65*, 226-256.

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

7. Haber, M., Maximum likelihood methods for linear and log linear models in categorical data. *Computational Statistics & Data Analysis*, 1985, *3*, 1-10.

8. Haberman, S.J., *Analysis of qualitative data: Introductory topics* (Vol 1). New York: Academic Press, 1978.

9. Koch, G.G., and Reinfurt, D.W., The analysis of categorical data from mixed models. *Biometrics, 1971, 27,* 157-173.

10. Kennedy, J.J., *Analyzing qualitative data: Introductory log-linear analysis for behavioral research.* New York: Praeger Publishers, 1983.