Multiple Regression Analysis: Applications in the Health Sciences.
D. Herbert and R. Myers (eds). New York: The American Institute of Physics (1986), 84-123.

REGRESSION METHODS FOR BINOMIAL AND POISSON DISTRIBUTED DATA

E. L. Frome*

Mathematics and Statistics Research Section
Engineering Physics and Mathematics Division
Oak Ridge National Laboratory
Oak Ridge, TN 37830

ABSTRACT

Models are considered in which a rate or probability can be represented by a regression function that describes the relation between the predictor variables and the unknown parameters. Estimates of the parameters can be obtained by means of iteratively reweighted least square (IRLS). When the dependent variable is a count that follows either the Poisson or binomial distribution, the IRLS algorithm is equivalent to using the method of scoring to obtain maximum likelihood (ML) estimates. This general least squares regression approach includes linear, generalized linear, and intrinsically nonlinear regression functions. Standard statistical packages that support IRLS can be used to obtain ML estimates, their asymptotic covariance matrix, and diagnostic measures that can be used to aid the analyst in detecting outlying responses and extreme points in the model space. The results of fitting several different models to the same data set can be summarized in an ANOVA-like table using the deviance as a measure of residual variation. Five examples using data from both designed experiments and observational studies are presented to illustrate the utility of Poisson and binomial regression analysis.

1. INTRODUCTION

1.1 *Notation and Terminology*

In this paper we assume that data have been obtained on each of N (*observational or experimental*) *units* . The data for the *ith* unit consist of the following:

y_i -- the observed value of the dependent variable,
c_i -- the "sample size", and
$X_i = (x_{i1}, x_{i2}, \ldots, x_{im})$ -- a row vector of covariates,

where x_{i1} is value of the first covariate, etc. If the values of X_i and c_i are determined in advance by the investigator (and randomization is used) then the results are from a *designed experiment*. In many situations, primarily in human populations, the investigator is restricted to observing the

*Edward L. Frome is on the research staff, Mathematics and Statistics Research Section, Engineering Physics and Mathematics Division of Oak Ridge National Laboratory, P. O. Box Y, Oak Ridge, TN 37830. This research was supported by Contract W-7405-eng-26 between U.S. Department of Energy and Union Carbide Corporation, and by Contract No. DE-AC05-760R00033 between Oak Ridge Associated Universities and the U. S. Department of Energy, Office of Health and Environmental Research. The author is grateful to the staff of the Center for Epidemiologic Research, ORAU for offering helpful suggestions, and to Donna Poole for help with manuscript preparation. By acceptance of this article, the publisher or recipient acknowledges the U. S. Government's right to retain a nonexclusive, royalty-free license in and to any copyright covering the article.

values of the covariates, and we refer to such situation as *observational studies* (see Cochran, 1983). In both situations one or more of the covariates may correspond to "causal factors" (i.e. treatments, procedures, or programs) that are of primary interest, while other x variables may affect the variation in the dependent variable and are included so that their influence can be "adjusted for" in the statistical analysis. We assume that the y_i are counts, and use either the Poisson or binomial distribution as a model for the variation in the dependent variable. The term *covariate* (or x-variable) will be used as a generic term for what are sometimes called independent variables, predictor variables, explanatory variables, or stimulus variables. Covariates may be either quantitative or qualitative in nature, and we use the term *factor* to refer to qualitative covariates. A factor can take a limited number of values called *levels*. For example, if A is a factor with k levels, then these can be coded using the integers $1,2,\ldots,k$. The *actual* levels of the factor may be either qualitative or quantitative. The levels of a factor are used to generate "dummy variables" (indicator variables) in the covariate vector, i.e.

$$x_{i1} = \begin{cases} 1 \text{ if } \textit{level } 1 \textit{ of factor } A \textit{ is present on unit } i, \\ 0 \textit{ otherwise} \end{cases}.$$

This implies that there will be a parameter associated with each level of the factor.

In situations that involve a linear structure (see below) a useful notation has been developed (Wilkinson and Rogers, 1973) and adapted for use in computer programs (see e.g. GLIM-3, Baker and Nelder, 1978). In agriculturial experiments, for examples, blocks and varieties are examples of factors. In observational studies factors might include sex, socioeconomic status, geographic region, etc. Table 1 contains further examples of covariates, as well as a description of the dependent variable y and sample size for each example that we will consider in Section 2 and 3.

Table 1. List of Examples of Data Used To Illustrate Poisson and Binomial Regression

Example	Dependent Variable y	Sample Size c	Distribution	Covariates x
1 and 2	number of chromosome aberrations	total lymphocytes scored (unit = 100)	Poisson	radiation dose, exposure rate
3	number of lung cancer deaths	man-years (units = 10^5)	Poisson	smoking, age
4	number of damaged lymphoblasts	total lymphoblasts examined	binomial	streptonigrin dose
5	number of mice with liver neoplasms	total mice examined	binomial	2-AAF exposure, time

Assuming that we have data y_i, c_i, X_i, $i=1,...,N$ on each of N units, the problem of interest is to determine if there is any systematic relation between the dependent variable and the covariates. We assume that the expected value of y_i is given by

$$\mu_i = c_i \lambda(X_i,\beta),$$

where $\lambda(X,\beta)$ is a known function, and $\beta = (\beta_1,..., \beta_p)'$ is a p-dimensional vector of unknown parameters. The regression function will in general be nonlinear in the unknown parameters, and hopefully p (the number of parameters) is much less than N. A special case of interst occurs when the regression function can be expressed as a linear combination of the covariates and the parameters, i.e.

$$\lambda(X_i,\beta) = X_i\beta = \sum_{j=1}^{p} \beta_j x_{ij}.$$

This is referred to as a (multiple) linear regression (function), and is most often encountered in the context of the "classical linear model". There is an extensive literature on the classical linear model (see Draper and Smith, 1981), and these methods are most appropriate when the y_i are continuous and follow the Normal distribution with constant variance. When the dependent variable is a count that may follow the Poisson or binomial distribution the assumptions of constant variance is clearly inappropriate, and the linear regression function is of limital value. A more general class of regression functions that involves a linear structure has developed for count data. A *generalized linear function* (GLF) is defined by

$$\lambda(X_i,\beta) = G(X_i\beta),$$

where $G(\quad)$ is a monotonic differentiable function. A widely used GLF for Poisson data is the *product model* $\lambda(X_i,\beta) = \exp(X_i\beta)$, which is also called a log-linear model. A useful GLF for binomial data is $\lambda(X_i,\beta) = \exp(X_i\beta)/\left[1 + \exp(X_i\beta)\right]$. The use of GLFs for Poisson and binomial data has been discussed by Nelder and Wedderburn (1972), and the statistical program GLIM (1977) has been developed to facilitate data analysis using GLFs. Nelder and Wedderburn use the term "link function" for the inverse of the function G, i.e. for the product model the link function is log (λ) which explains why this is also called a log-linear model.

In summary, we use the term *Poisson (binomial) regression model* to refer to a situation where

i) the y_i are independent and follow the Poisson (binomial) distribution, and

ii) the regression function $\lambda(X_i,\beta)$ is known.

Then given the data we want to obtain estimates of the β_j, their standard deviations, and evaluate the "goodness of fit" of the regression model. In what follows we show that estimates of the β_j can be obtained using a general regression model (1.1), from which certain generalized least squares equations (1.3) are derived. A root of the system of equations can be obtained using an itertively reweighted least squares procedure which is equivalent to maximum likelihood estimation under the Poisson or binomial assumption.

1.2 *General Regression Model*

Consider the general regression model

$$E(y_i) = \mu(X_i,\beta) = c_i\lambda(X_i,\beta)$$

$$Var(y_i) = \sigma^2 V(\mu_i) \quad , \quad i = 1, ..., N , \tag{1.1}$$

where y_i is the response for the *ith* unit, c_i denotes the "sample size", and V( ) is a known function that may depend on μ_i. We assume that the y_i are uncorrelated, that $\lambda(X,\beta)$ is a known function of the m dimensional row vector of covariates $X_i = (x_{i1}, ..., x_{im})$, and the p unknown parameters $\beta = (\beta_1, ..., \beta_p)'$. The *regression function* $\lambda(X,\beta)$ will in general be nonlinear (with respect to the parameters) and we assume that it is a differentiable function of β. The regression function relates the expected value of the dependent variable y to the covariates and the parameters, and given the data $\{y_i,c_i,X_i , i = 1, ..., n \}$ we want to estimate the unknown parameters.

The most widely used methods of estimation have been developed using either the maximum likelihood (ML) or the least squares (LS) principle. The assumptions underlying these principles of estimation and the properties of the estimators are well known (see e. g. Kendall and Stuart, 1946 or Rao, 1965). Following the least squares principle we define the weighted sum of squares

$$S(\beta) = \sum_{i=1}^{N} w_i \, [\bar{y}_i - \lambda(X_i,\beta)]^2, \tag{1.2}$$

where $\bar{y}_i = y_i/c_i$, and w_i is a positive weight that is inversely proportional to the variance of $\bar{y}_i$. The least squares estimates are obtained by solving the p dimensional system of equations

$$\mathbf{P}' \, \mathbf{W} \, [\, \bar{\mathbf{y}} - \lambda(\beta)] = 0, \tag{1.3}$$

where $\mathbf{W} = diag[w_i]$, $\mathbf{P}$ is an N by p matrix of partial derivatives with elements $p_{ij} = \partial\lambda(X_i,\beta)/\partial\beta_j$, $\bar{\mathbf{y}} = (\bar{y}_1, ..., \bar{y}_N)'$, and $\lambda(\beta) = (\lambda(X_1,\beta), ..., \lambda(X_N,\beta))'$. A solution of the generalized least squares equations (1.3) can be obtained using an iterative reweighted least squares (IRLS) procedure (see the Appendix).

It is well known that when the y_i follow the Normal distribution, the IRLS procedure will yield ML estimates for the general regression mode (1.1). Moore and Seigler (1967) showed the equivalence of ML and LS for the binomial distribution, and the equivalence of the ML and LS estimation procedure for Poission distributed data was demonstrated by Frome and Beauchamp (1968), Frome (1972), and Frome, Kutner, and Beauchamp (1973). The equivalence of IRLS and ML for certain generalized linear functions for y in the exponential family (this includes Normal, gamma, Poisson, and binomial) was established by Nelder and Wedderburn (1972). Charnes, Frome, and Yu (1976) extended this result to show that ML and IRLS are equivalent for general nonlinear regression functions when the dependent variable is in the exponential family.

Regression methods based on the classical linear model assume that the y_i are continuous variables, that the regression function is linear in the β_j – i.e. $\lambda(X_i,\beta) = X_i\beta = \sum_j \beta_j x_{ij}$, and that the y_i have equal variances and are uncorrelated. The assumption of Normality of the distribution of y leads to an exact small-sample theory. The theory of least squares in large samples can, however, be developed using only the first- and second-moment assumptions. The classical linear model is most appropriate when the dependent variable is a continuous quantity that can take on values on the entire real line. In practice it is often used to model data that are continuous and nonnegative (e.g. weights, concentrations, etc.), provided the mean values are far from zero. In some situations the use of log y instead of y may justify the assumption of constant variance or Normality.

1.3 *Poisson and Binomial Regression Models*

There are many situations in which the dependent variable is a count and the investigator is interesting in evaluating the effect of one or more covariates on the response. The two most widely used probability models for discrete data are the Poisson and the binomial distribution, and we shall limit our discussion to these two distributions.

The Poission distribution has been widely used as a model for certain types of discrete data. Haight (1967) states that the Poisson is second in importance to the Normal distribution, both in terms of abstract theory and breadth of application. The Poisson distribution has only recently been considered in the context of regression analysis (see Frome, 1972, Frome, Kutner and Beauchamp, 1973, Kock, Athinson, and Stokes, 1984). The dependent variable y is typically a count that is made with respect to some reference quantity c that is a measure of the size of the sample. Examples are bacteria per unit volume of suspension, number of accidents per unit time, and number of abnormal chromosome per lymphocyte. Cochran (1940) was the first to propose the use of Poisson regression in the context of the analysis of variance for designed experiments. Cochran used a linear regression function $\lambda(X_i,\beta) = \sum_j \beta_j x_{ij}$ to study the effect of several factors on the yield of wireworms (per acre). He found that the exact solution to the likelihood equations was too complicated for frequent use, and proposed $\lambda(X,\beta) = (X\beta)^2$ as a more

practical alternative. This "square-root-linear" model results in some simplification (an iterative procedure is still required) of the likelihood equations. Cochran [1940] also proposed the product (log-linear) model i.e. $\lambda(X,\beta) = \exp(X_i\beta)$, for Poisson distributed data.

Another important area where Poisson regression models are used is in the analysis of data on rates and survivorship in medical and epidemiologic studies. For the *ith* subgroup the dependent variable y_i is a count (e.g. lung cancer deaths), c_i is the total follow-up time, and X_i is a vector of covariates that describes the *ith* subgroup -- see Holford (1980), Frome (1983), and Breslow *et al* (1983).

For a general Poisson regression model y_i denotes the observed count for the *ith* set of covariate vector X_i, c_i is the "size" of the *ith* unit and $\bar{y}_i = y_i/c_i$ denotes the observed rate (i.e. bacteria per liter, failures per hour, etc.). The expected number of event is $\mu_i = c_i\ \lambda(X_i,\beta)$, and the regression function $\lambda(X,\beta)$ can be interpreted in a general sense as the expected rate. Under the Poisson assumption the $Var(y_i) = c_i\lambda_i$, and consequently the "Poisson weights" in (1.1) are defined by $w_i = c_i/\lambda_i$. Examples that illustrate the use of linear, log-linear, and intrinsically nonlinear models are given in Section 2.

The binomial distribution--which is one of the oldest to have been studied (see Johnson and Kotz, (1970) Chap. 3, -- has two parameters, say λ (the probability of success on a given trial) and c (the number of trials). The dependent variable y is the number of positive responses that are observed in the c trials. In a regression context we assume that c_i individuals are observed for the *ith* set of experimental conditions X_i and y_i show a positive response with expected value $c_i\lambda(X_i,\beta)$, where $\lambda(X_i,\beta)$ is the probability of response for each individual. For example, in a bioassay c_i would be the number of animals and y_i would be the number that respond to the *ith* set of experimental conditions that are defined by the covariate vector X_i. In this situation the regression function $\lambda(X_i,\beta)$ represents the probability of observing the response for each animal in the *ith* group, and $\bar{y}_i = y_i/c_i$ denotes the proportion that respond. The response must lie in the range $0 \leqslant y_i \leqslant c_i$, and the regression function must satisfy $0 \leqslant \lambda(X_i,\beta) \leqslant 1$. Under the binomial assumption the variance (see eq. 1.1) is given by $Var(y_i) = c_i\ \lambda_i\ (1-\lambda_i)$, and consequently the "binomial weights" in (1.2) are defined to be $w_i = c_i/[\lambda_i(1-\lambda_i)]$. The best known binomial regression function is encounted in probit analysis (Bliss, 1935). The probit regression function is often defined by $\lambda(X_i,\beta) = \Phi(\alpha+\beta d_i)$ where Φ is the Normal cumulative distribution function, $X_i = (1,d_i)$, and $\beta = (\alpha,\beta)'$. Typically, d_i is the dose (or logarithm of the dose) of a stimulant or toxin, y_i is the number of positive responses in the *ith* exposure group, and c_i is the number of trials. The role of the logistic function in the analysis of "binary" or "quantal" response data has been considered by Cox (1970). Note that if $c_i = 1$ for all values of i then the possible values of y_i are 0 and 1. In this case

$$E(Y_i) = prob\ (Y_i=1) = \lambda(X_i,\beta) = \exp(X_i\beta)/(1+\exp(X_i\beta))$$

The more general case with all c_i greater than one is sometimes referred to as "grouped" binary data.

Further results and an extensive bibliography concerning the use of GLFs in the analysis of discrete data are given by McCullagh and Nelder (1983). This includes topics such as log-linear models for contingency tables (see also Bishop *et al*, and Huberman, 1974) and the connection between log-linear and multinomial response models (see also Palmgren, 1981).

2. POISSON REGRESSION

In this section three examples will be presented to illustrate various aspects of Poisson regression analysis. Example 1 will be used to illustrate simple linear regression analysis, multiple linear regression, and the analysis of variance for Poisson distributed data. In Example 1 there are parallel counts (i.e. replication) for each set of experimental conditions, and consequently we can test for "lack of fit" of the regression function and for heterogenity of variance. Example 2 will be

used to further illustrate linear regression, and to introduce nonlinear regression and regression diagnostics for Poisson data. In Example 3 we will consider a log-linear regression function and a nonlinear regression function for data obtained from an observational study.

2.1 *Simple Linear Regression with Parallel Counts*

Example 1: Ir-192 Dose-Response Curve. In cytogenetic dosimetry *in vitro* dose response curves are used to describe the relation between the yield of dicentric chromosome aberrations and radiation dose. Let y_{jk} denote the observed dicentric yield for the kth parallel count of human lymphocytes exposed to d_j grays -- see DuFrain *et al* (1980), Frome and DuFrain (1983). The data in Table 2a provide a numeric example of an *in vitro* dose response curve for Ir 192 and are shown graphically in Figure 1a. As a matter of convenience we define c_{jk} as the total cells scored in units of 100 cells so that the regression function $\lambda(d,\beta)$ represents the dicentric yield per 100 cells.

The dicentric yields follow the Poisson distribution and we first assume that $\lambda(d,\beta) = \beta d$, i.e. the dicentric yield can be represented by a straight line that passes through the origin. In this case the ML estimate is

$$\beta^* = \frac{\sum_j \sum_k y_{jk}}{\sum_j \sum_k c_{jk}\, d_j} = \frac{\sum_j c_{j.}\, \bar{y}_j}{\sum_j c_{j.}\, d_j}$$

where $c_{j.} = \sum_k c_{jk}$. This is identical to the least squares estimate with $w_j = c_{j.}/d_j$. The ML estimate is $\beta^* = 21.4$, and the deviance is 79.26 with 19 df (see Table 2c). Standardized residuals for this model

$$u_{jk} = (y_{jk} - c_{jk}\, \beta^*\, d_j)/(c_{jk}\, \beta^*\, d_j)^{1/2}$$

are listed in Table 2b and are shown graphically in Figure 1b. The residuals suggest that a dose-squared term should be included in the model, i.e. $\lambda(d) = \beta_1 d + \beta_2 d^2$ (see Section 2.2.1)

2.2 *Generalized Linear Regression*

There are two general forms of the regression function that are of practical interest for Poisson distributed data. They are the "multiple linear"

$$\lambda(X_i\beta) = \sum_j \beta_j x_{ij},$$

and the "log-linear"

$$\lambda(X_i\beta) = \exp(\sum_j \beta_j x_{ij}),$$

regression functions. A third regression function of historical interest is

$$\lambda(X_i\beta) = (\sum_j \beta_j x_{ij})^2,$$

and this might be called "square-root-linear". All of these GLFs can be viewed as a special case of

$$G(X_i\beta) = (\sum_j \beta_j x_{ij})^q \ , \ q \neq 0$$

$$G(X_i\beta) = \exp(\sum_j \beta_j x_{ij}) \ , \ q = 0 \qquad (2.1)$$

Nelder and Wedderburn (1972) use the term "link function" to describe the inverse of the regression function for GLFs, i.e. they call the exponential regression function $G(X\beta) = \exp(X\beta)$ a log-linear model because log $G(X\beta)$ is linear in the parameters. In Section 2.2.1 we will consider several multiple linear regression functions for the Ir-192 data from Example 1. Then, in Example 2, we will consider an experiment in which two independent variables, radiation dose and exposure rate, for Cs-137 gamma rays are of interest. Linear regression models will be used to investigate the effect of dose and dose rate on the frequency of chromosome abnormalities. In Section 2.3 we will consider a theoretical model derived from the

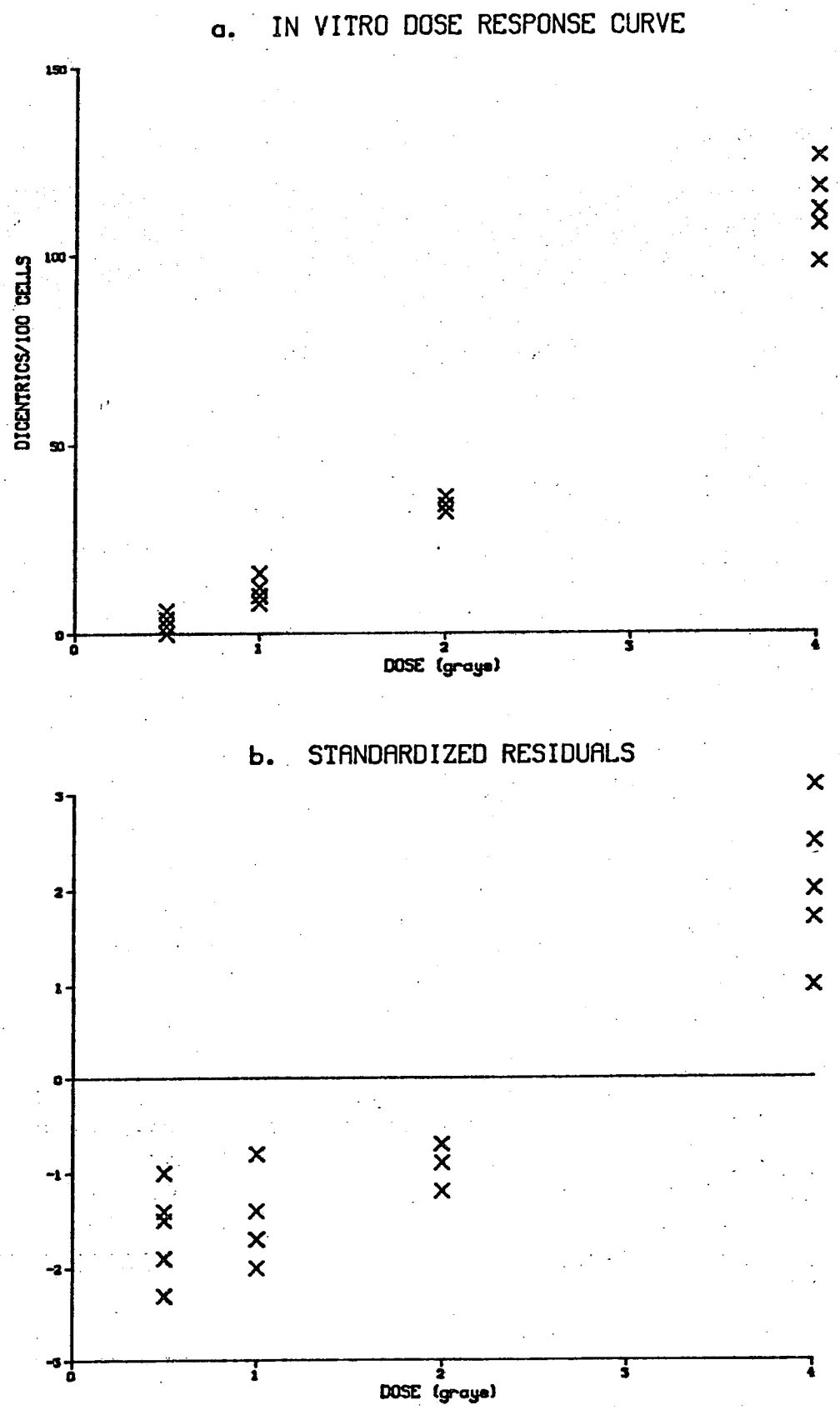

Figure 1. IRIDIUM GAMMA RAY DOSE-RESPONSE DATA.
(See Section 2.1 and Table 2a)

theory of dual radiation action (DRA) that is nonlinear in the parameters. The Cs-137 data from Example 2 is used to illustrate Poisson regression analysis for the intrinsically nonlinear DRA model.

In Section 2.2.2 an example of Poisson regression analysis for data from an observational study will be considered. The dependent variable is the number of lung cancer deaths and the covariates are age and exposure rate for individuals who were regular cigarette smokers. A log-linear model is used to obtain age adjusted estimates of "smoking effects". In Section 2.3 we will consider a nonlinear model (derived from the multi-stage theory of carcinogenesis) to further illustrate Poisson regression analysis for an intrinsically nonlinear regression function.

2.2.1 Multiple-Linear Regression

Example 1 - continued. In Section 2.1 we found that a simple one parameter model was inadequate, and suggested that a dose-squared term should be added to the regression function, i.e.

$$\lambda(X_i,\beta) = X_i\beta = (x_{i1},x_{i2}) \begin{bmatrix}\beta_1\\ \beta_2\end{bmatrix}$$

where $x_{i1}=d$ and $x_{i2}=d^2$. The deviance for this model is 12.76 with 18 degrees of freedom (df) – see line 6 in Table 2c. Table 2c contains the value of the deviance for all possible multiple linear regression models of the form

$$\lambda(d) = \beta_1 + \beta_2 d + \beta_3 d^2.$$

The next to the last line in Table 2c labelled "Dose Groups" corresponds to the "pure error sum of squares" in an ANOVA table for the standard linear models. The rows in the model matrix are based on indicator variables, i.e. for $j=1,\ldots,4$

$$x_{ij} = \begin{cases} 1 \text{ if } y_i \text{ is in dose group } j, \\ 0 \text{ otherwise.} \end{cases}$$

The ML estimates are of course

$$\beta_j^* = \frac{\sum y_i}{\sum c_j} = \bar{y}_j,$$

where the summation is over units in the jth dose group (i.e. $\sum_{k=1}^{n_j} y_{jk}$ if the double subscript notation of Section 2.1 is used). The deviance for this model is important since it provides a test of the assumption of Poisson variation. In this example the value of the deviance is 11.29 and is compared with the chi-square distribution with 16 df, indicating that the assumption of Poisson variation should not be rejected. Another measure of residual variation that can be used for this purpose is Fisher's index of dispersion

$$\sum_{j=1}^{N} \sum_{k=1}^{n_j} (y_{jk} - \bar{y}_j)^2 / \bar{y}_j \quad .$$

The value of this pooled within dose group index of dispersion is 9.51 with $\sum_j(n_j-1) = 16\ df$. Both of these lack of fit statistics can be viewed as a sum of squares of standardized residuals within experimental units.

If the Poisson assumption is not rejected then test statistics obtained from the Poisson ANOVA table can be compared with the appropriate chi-square distribution. For example, in Table 2c this would lead us to reject the first two models. The Poisson ANOVA in Table 2c has been given for illustrative purposes, and might be appropriate in situations of an exploratory nature where little was known about the dose-response relation. A more appropriate Poisson ANOVA table from a biologic point of view is given in Table 2d. This table indicates that the linear-quadratic model $\lambda(d) = \beta_1 d + \beta_2 d^2$ provides a better representation of these data than the linear model since the likelihood ratio statistic for $H_0{:}\beta_2 = 0$ is 66.6 with 1 df. This of course agrees with the "lack-of-fit" test on line 1 of Table 2d. The "lack-of-fit" test on line 4 (which is not significant) implies that the linear-quadric regression function not only provides a

significant improvement over the linear function, but that it cannot be rejected as a reasonable model for these data (see Table A.2 in the Appendix). Note that this is a true test for lack of fit of the regression function in the sense discussed by Draper and Smith (1966, Chap. 2).

The ML estimates of the parameters and their estimated standard deviations (in parenthesis) are $\beta_1^* = 3.59$ (1.68) and $\beta_2^* = 6.22$ (0.67). In summary, the IR-192 dose response data are best represented by the L-Q model, and the Poisson distribution provides an acceptable model for the variation in the dicentric counts.

Table 2 Poisson Regression Example

a) In-Vitro Dose-Response Data for Human Lymphocytes Exposed to 192 Ir Gamma Radiation

Dose (Grays)	Dicentric Chromosomes Per 50 Cells Scores							n_j	Total Dicentrics	Total Cells Scored
.5	0	1	0	2	1	3	2	7	9	350
1.0	5	6	5	4	8			5	28	250
2.0	16	17	18					3	51	150
4.0	49	59	54	56	63			5	281	250

Notes: 1) See DeFrain *et al* (1980) for further discussions.

2) $E(y_{jk}) = c_i\lambda(X_i,\beta) = c_{jk}\ \beta\ d_{jk}$

3) $ML\ \ Estimate\beta^* = \sum_j \sum_k y_{jk} \ / \ \sum_j \sum_k c_{jk}\ d_{jk}$

b) Residuals for Simple Linear Regression for Dicentric Data in Table 2.1a

Dose d	Standardized Residual-- $u_{jk} = (y_{jk} - y^*_{jk})/(y^*_{jk})^{1/2}$						
0.5	-2.3	-1.9	-2.3	-1.4	-1.9	-1.0	-1.5
1.0	-1.7	-1.4	-1.7	-2.0	-0.8		
2.0	-1.2	-0.9	-0.7				
4.0	1.0	2.5	1.7	2.0	3.1		

c) Poisson ANOVA For Data in Table 2.1(a) All Possible Models

Model	Number of Parameters	Unexplained Variation+	df
const	1	514.2	19
d	1	79.26	19
d^2	1	18.05	19
const+d	2	23.13	18
const+d^2	2	14.58	18
d + d^2	2	12.67	18
const+d+d^2	3	11.34	17
Dose Groups	4	11.29	16
Complete	20	0.0	0

d) Poisson ANOVA For Models of Cytogenetic Interest

Model	Number of Parameters	Unexplained Variation+	df	Likelihood Ratio Statistic For Lack-of-Fit	df
d	1	79.26	19		
				66.59	1
*d+d^2	2	12.67	18		
				1.38	2
Within Dose Groups	4	11.29	16		
Complete	20				

*$\lambda(d) = 3.6d + 6.2d^2$

+ The deviance $D(\beta^*)$ is used as a measure of unexplained (i.e. residual) variation (see Appendix)

Example 2: Caesium Dose-Response Curves. The data in Table 3 (Purrott and Reeder, 1976) were obtained from an experiment (using gamma radiation from a caesium-137 source) that was designed to investigate the effect of dose rate on dicentric yield. The observed rate ($\bar{y} = y/c$) of dicentric induction (per 100 cells scored) is shown graphically in Figure 2 for each of the nine exposure rate groups. According to theoretical predictions from microdosimetry, a quadratic dose-response relation is predicted for low LET radiation, i.e. dicentric frequency is equal to $\alpha d + \beta d^2$ where d is radiation dose. From a biological point of view the two coefficients are thought of as corresponding to two different physical events. The linear term describes the induction of dicentrics by a single ionization or track, and the dose squared term which describes the induction of dicentrics by two different ionizations or tracks. Thus, the two break asymmetic exchange (dicentric) frequency is believed to be the result of these two phenomena, and is described by a second degree polynomial in dose. The validity of the quadratic model is predicated on the assumption that the absorbed dose is delivered to a "critical site" in a short period of time, i.e. at a high dose rate.

The purpose of the study by Purrott and Reeder was to test the hypothesis that the effect of decreasing the dose-rate would be to decrease the contribution of the dose-squared term, without

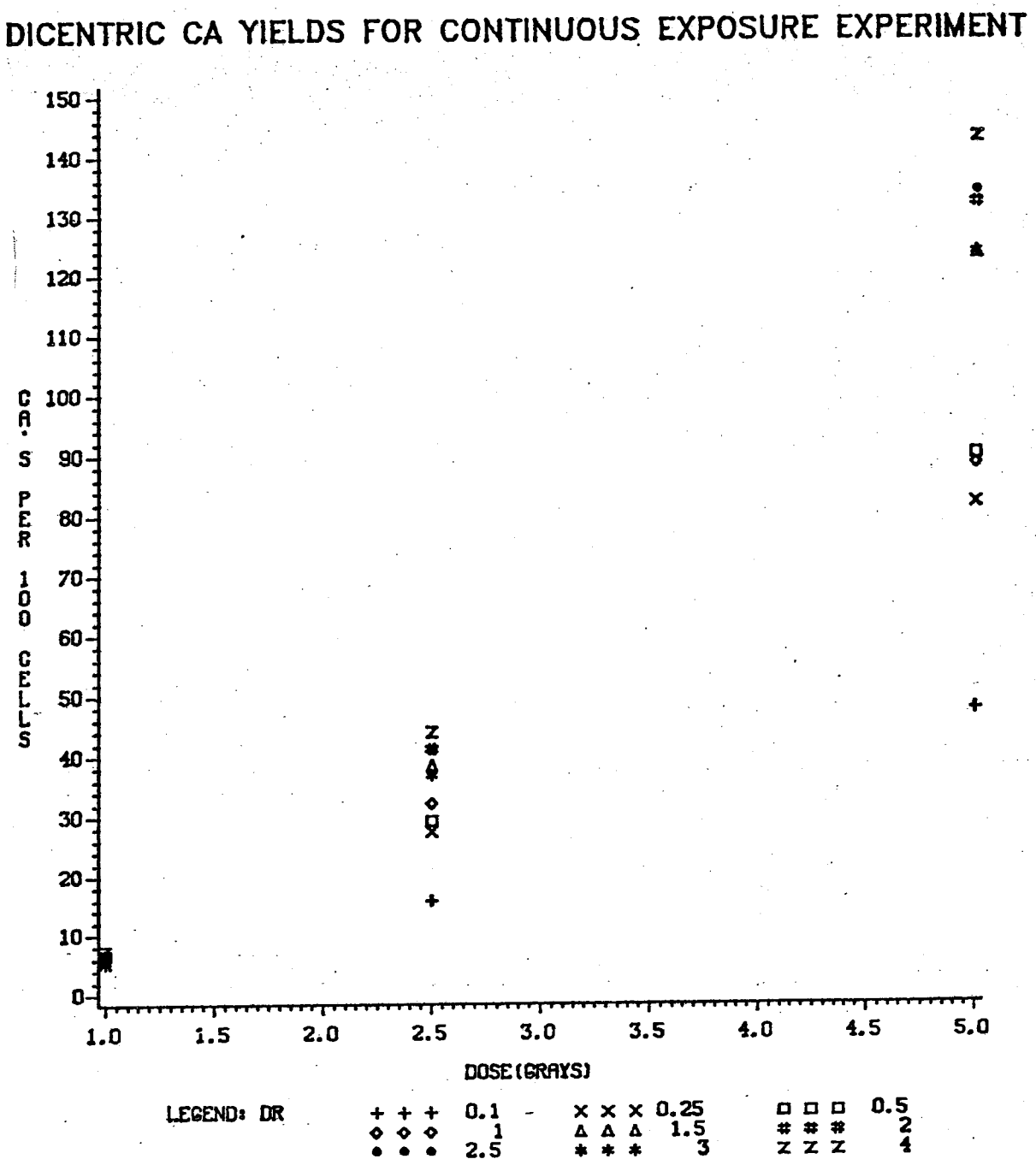

Figure 2. CYTOGENETIC DOSE RESPONSE DATA.
(See Section 2.2 and Table 3)

changing the linear term. Model 4 (see Table 4) corresponds to the most general case in which both the linear and quadratic coefficients are allowed to vary with dose rate, i.e. $\lambda_{jk} = \alpha_j d_k + \beta_j d_k^2$, where j identifies the dose rate group. For each of the models in Table 4 the regression function $\lambda(X,\beta)$ is linear in the parameters, and the procedure described in the Appendix was used to obtain the Poisson ANOVA. A test statistic for the hypothesis $\beta_1 = \beta_2 = \cdots = \beta_g$ is obtained using the difference of the deviance $D[y,\mu^*(2)] - D[y,\mu^*(3)] = 206.48$. This test statistic has an asymptotic chi-squared distribution with 8 df, if the more restrictive hypothesis is true. Consequently, we reject the hypothesis that the coefficient of the quadratic term is independent of dose rate. An alternative approach is to test for "lack of fit" of model 3. The deviance for this model is 21.52 with 17 df indicating that model 3 cannot be rejected.

It the ML estimates of the quadratic coefficients obtained from model 3 are plotted against the log of the dose rate it appears that the β_j^* increase linearly with log dose rate, and this can be described by the following regression model

$$\lambda_{jk} = \alpha d_k + \left[(\theta_1 + \theta_2 \log_{10}(r_j)\right] d_k^2. \tag{2.2}$$

The *ith* row of the model matrix for this *ad hoc* model is $X_i = (d_i, d_i^2, d_i^2 \log_{10} r_j)$. The ML estimates and estimated standard errors for this model are given in Table 5. The value of the deviance for the model is 29.95 with 24 df, indicating that this *ad hoc* model cannot be rejected for these data. This model provides a good description of effect of dose rate on dicentric yield, i.e., the quadratic component increase with the log of dose rate, and the linear component is independent of dose rate.

The results of this initial analysis (using linear regression functions to describe the effect of dose and dose-rate on dicentric yield) represents a straightforward extension of results from the standard linear model to Poisson distributed data. Although this analysis is technically correct we decided to reject this approach as being both inappropriate and misleading on biologic grounds (see Frome and DuFrain, 1983). A more appropriate analysis that utilizes a nonlinear model that was derived from the theory of dual radiation action (Kellerer and Rossi, 1972) is given in Section 2.3 of this paper.

Table 3. Cytogenetic Dose-Response Curve Data for Continuous Exposure Experiment

Dose Rate G/hr	Dose (Grays) 1.0 c	1.0 y	2.5 c	2.5 y	5.0 c	5.0 y
.1	4.78	25	3.28	52	2.10	100
.25	19.07	102	1.85	51	1.38	113
.5	22.58	149	3.42	100	1.60	144
1.0	23.29	160	3.10	100	1.20	106
1.5	12.38	75	2.78	107	.90	111
2.0	14.91	100	2.59	107	1.00	132
2.5	15.18	99	2.49	102	3.13	419
3.0	7.64	50	2.98	110	1.82	225
4.0	13.67	100	2.43	107	1.44	206

NOTE: y = number of dicentrics, c = cells scored (100s)

Source: Purrott and Reeder (1976)

Table 4. Poisson ANOVA for Cytogenetic Data in Table 3

	Regression Model	Number of Parameters	Deviance*	df
1	αd_i	1	1075.30	26
2	$\alpha d_i + \beta d_i^2$	2	228.00	25
3	$\alpha d_i + \beta_j d_i^2$	10	21.52	17
4	$\alpha_j d_i + \beta_j d_i^2$	18	11.10	9
5	Complete	27	0.0	0

* See Footnote to Table 1d

Table 5. Maximum Likelihood Estimates for Ad Hoc Model for Dose-Response Curve Data in Table 3

Parameter	Estimate	Standard Deviation
α	2.86	.305
θ_1	3.80	.141
θ_2	2.26	.144

2.2.2 Log-Linear Regression

Incidence or mortality data obtained from epidemiologic follow-up studies are often expressed as covariate stratum-specific rates, where the covariate may be age or some other presumed confounding factor. Poisson regression provides a general approach to the study of the effect of one or more covariates on disease rates -- see Frome (1983), Frome and Checkoway (1984). the attractive features of the Poisson regression approach are that summary estimates of relative risk can be obtained, an evaluation of the presence and nature of interaction is part of the analysis, and the modeling of disease rates is facilitated. Poisson regression methods are especially appropriate in follow-up studies where time-based denominators (person-years) are used to obtain disease rates in a life table type of format (Frome, 1983), or when the outcome of interest is rare so that the Poisson approximation to the binomial distribution can be used (Gart, 1978).

Example 3: Lung Cancer Mortality The data in Table 6 were obtained by Kahn (1966) in a study of lung cancer mortality in relation to cigarette smoking. The dependent variable y_{jk} is the number of lung cancer deaths for the *jth* level of the potential confounding variable (age) and the *kth* level of the "risk factor", cigarette consumption. The c_{jk} are the person-years (pys) at risk (in units of 10^5 pys) and consequently the $\bar{y}_{jk}$ are lung cancer death rates per 10^5 pys. The y_{jk} are assumed to follow the Poisson distribution with expectation $\mu_{jk} = c_{jk}\ \lambda_{jk}$, where λ_{jk} denotes the underlying regression function. If the covariate stratum-specific RRs are constant within each risk group, then $\lambda_{jk} = \lambda_j\ \phi_k$ where:

λ_j denotes the rate for the *jth* stratum level, and

ϕ_k is the summary risk ratio for risk group k $(k>1)$ and $\phi_1=1$.

This is referred to as the *product model,* and for estimation purposes it can be expresed as a GLF (where $j=1,...,J$ and $k=1,...,K$)

$$\lambda_{jk} = \exp(\alpha_j + \delta_k) = \exp(X_i\beta). \tag{2.3}$$

where $\alpha_j = \log\lambda_j$ $(j=1,...,J)$ and $\delta_k=\log(\phi_k)(k=2,...,K)$.

We have assumed that risk group 1 is the reference, or non-exposed group. Consequently, the α_j correspond to the natural logarithms of the stratum specific rates in the reference group, while the δ_k are the logarithms of the summary RR for group k (with group 1 as the reference group). In (2.3) X_i is a $p = J+K-1$ dimensional row vector of indicator variables for the *ith* cell in the table, and $\beta = (\alpha_1,...,\alpha_J,\delta_2,...,\delta_K)'$ is the p-dimensional column vector of unknown parameters. If the *ith* cell of the table corresponds to row j and column k, then the components of X_i, $(i=1,...,JK)$ can be defined as follows:

$$x_{im} = 1 \text{ if } m=j; x_{im} = 1 \text{ if } k>1 \textit{ amd } m = J+k-1;$$

$$\text{for } m=1,...,J+K-1, \text{ and } x_{im} = 0 \textit{ otherwise.}$$

When $c_{jk} > 0$ for all j and k, this situation is equivalent to a full rank parameterization of the design matrix for a two factor fixed effects ANOVA model. In practice it is not necessary to generate this matrix because its structure is implied by the levels of the factors. The ML estimates of the parameters for the product model (2.3) for the data in Table 6 are given in Table 7. The deviance is 12.5 with 10 df indicating that the product model provides a good description of these data. The ML estimates of the λ_js and ϕ_ks are given in the last row and column of Table 6b, respectively.

If the risk factor (smoking) is not important, i.e. $\delta_k = 0$(for$k=2,...,K$), then $\lambda_{jk} = \lambda_j$ and the ML estimates are $\lambda_j^* = \sum_k y_{jk}/\sum_k c_{jk}$. The deviance for this model (see line 3 Table 8) is 1037 with 25 df, and the likelihood ratio statistic (obtained by subtracting line 4 from line 3 in Table 8) for the hypothesis $\delta_2 = ... = \delta_K = 0$ is 1024.5 with 5 df indicating that the risk ratios are highly significant. Frome and Checkoway (1984) have shown that when the product model provides a good fit (as it does in this example) the ϕ_k^* can be interpreted as estimates of standardized risk ratios (SRR). The SRR for risk group k (with $k=1$ for the reference group) is defined by Miettien (1974) as follows:

$$SRR_k = (\textstyle\sum_j w_j \bar{y}_{jk})/(\sum_j w_j \lambda_j),$$

where the w_j are standard population weights. If the product model provides a good fit (as indicating by the deviance), the $\bar{y}_{jk}$ in the above definition can be replaced by their ML estimates $\lambda_{jk}^* = \lambda_j^* \phi_k^*$ to obtain

$$SRR_k^* = \textstyle\sum_j w_j \lambda_j^* \phi_k^*/\sum_j w_j \lambda_j^* = \phi_k^*.$$

The ϕ_k^* are estimates of the SRRs with the non-exposed group as the referent group, and the choice of the standard population weights is unimportant.

Table 6. Lung Cancer Mortality According to Cigarette Consumption and Age

a) Number of Deaths and Person-Years (pys)

		Current Cigarette Smokers (cigarettes/day)					
			Age				
Group		Smokers	Occasional	1-9	10-20	21-39	40+
35-44	deaths	0	0	0	2	4	0
	pys	35164	3657	8063	59965	40643	3992
45-54	deaths	0	0	0	2	10	2
	pys	15134	1283	3129	16392	12839	1928
55-64	deaths	25	6	31	183	245	63
	pys	213858	14624	45217	151664	103020	19649
65-74	deaths	49	10	44	239	194	50
	pys	171211	10053	37130	101731	50045	8937
85-	deaths	4	1	5	15	7	3
	pys	8489	512	1923	3867	1273	232

Source: Kahn (1966)

b) Lung Cancer Deaths Rates (per 10^5 pys)

Age Group (midpoint)	Cigarettes/day 0	.5	5	15	30	45	Age Fit
40	0	0	0	3	10	0	0.4
50	0	0	0	12	78	104	3.2
60	12	41	69	121	238	321	14.0
70	29	99	118	235	388	559	25.5
80	47	195	260	389	550	1293	43.9
Smoking Effect	1.0	3.5	4.8	8.9	16.2	22.6	

*Age fit = $\exp(\alpha_j^*)$ and smoking effect = $\exp(\delta_k^*)$, where the α_j^* and δ_k^* are the ML estimates given in Table 7. The estimated lung cancer death rates per 100,000 man-years in Row j and Column k are

$$\lambda_{jk}^* = \textit{Age Fit} * \textit{Smoking Effect} = \exp(\alpha_j^* + \delta_k^*)$$

Table 7. ML Estimates of The Parameters for The Product Model (2.3) for the Lung Cancer Data in Table 6

j	α_j^*	St. Dev.	k	δ_k^*	St. Dev.
1	-0.82	.42	2	1.24	.27
2	1.18	.29	3	1.56	.16
3	2.64	.12	4	2.18	.12
4	3.24	.12	5	2.79	.12
5	3.78	.20	6	3.12	.15

Table 8. Poisson ANOVA Table for Lung Cancer Mortality Data in Table 6

Model	Log (λ)	No. of Parameters	Deviance*	d.f.
1. Minimal	α	1	1438.0	29
2. Smoking effect	$\alpha+\delta_k$	6	589.7	24
3. Age effect	α_j	5	1037.0	25
4. Age and smoking	$\alpha_j+\delta_k$	10	12.5	20
5. Complete		30	0	0

* See footnote to Table 1d

2.3 Nonlinear Regression

Example2: Caesium Dose-Response Curve (continued). The *ad hoc* mode (2.2) described in the previous section can be used as an empirical description of the cytogentic dose response relation for the experimental data in Table 3. The parameters in the ad hoc model do not have a clear interpretation in terms of the quantitative effects of ionizing radiation (see Frome and DuFrain, 1983). The dual radiation action (DRA) theory described by Kellerer and Rossi (1972) utilizes concepts from microdosimetry to provide a quantitative characterization of the effect of various temporal distributions of absorbed dose on the production of chromosome aberrations (CAs). It is postulated that elementary lesions are produced at a rate that is proportional to the square of the local energy concentration produced by charged particles in certain "critical sites". The form of the dose-effect model that is appropriate here (see Kellerer and Rossi, 1972, Section 5.4) is

$$\lambda(d,t) = \kappa[\gamma d+g(t,\tau)d^2], \tag{2.4}$$

where d denotes dose, t is time, and $\lambda(d,t)$ is the yield of elementary lesions. The parameter κ is a biophysical proportionality constant that reflects the target sensitivity for the biologic system (lymphocyte). The parameter γ depends on the radiation quality and can be related to the specific energy produced in a critical site by a single ionization. The linear term in (2.4) represents the effect due to intratrack interactions and the quadratic term represents the effect of intertrack interaction. The coefficient of the d^2 term is referred to as the "reduction factor", and assuming an exponential recovery process for continuous irradiation of duration t one obtains (see Lea, 1955)

$$g(t,\tau) = \frac{2\tau}{t} - \frac{2\tau^2}{t^2}(1-e^{-t/\tau}). \tag{2.5}$$

Using (2.5) in (2.4) leads to

$$\lambda(X_i,\beta) = \kappa\left\{\gamma d_i + \frac{2\tau}{t_i}\left[1 - \tau\{1-\exp(-t_i/\tau)\}/t_i\right] d_i^2\right\}, \tag{2.6}$$

where d is the absorbed dose and t is the duration of exposure at a constant dose rate.

The ML estimates of the parameters in (2.6) for the data in Table 3 were obtained using the IRLS procedure described in the Appendix. Since the DRA model is nonlinear in the parameters, the partial derivatives of (2.6) with respect to the parameters must be supplied. The ML estimates and their standard deviations are given in Table 9. The deviance for this model is 28.58 with 24 df (p = .236) indicating that the DRA model cannot be rejected. The standardized residuals in Table 10a are used to identify outlying observations, and in this example there is one large negative residual. The diagonal terms from the H matrix (see the Appendix) are given in Table 10b. There are several large h values (greater than $2p/n=0.22$) in column 3, and two of these are in the first two rows, i.e. the highest dose and the lowest exposure rates. The diagnostic quantities in Table 10 are shown graphically in Figure 3. Note that we have used scaled h values in this ($h^* = ph/n$) diagnostic plot so that two can be used as a cutpoint for large h^* values.

Table 9. ML Estimates for the DRA Model for the Cytogenetic Data in Table 3

Parameter	Estimate	Standard Deviation
κ	5.44	.208
γ	.269	.0677
τ	7.40	.857

Table 10. Regression Diagnostics for Data in Table 3 Using the Nonlinear DRA Model (2.6)

(a) Standardized Residuals $u_i = (y_i - \mu_i^*)/\mu_i^{*1/2}$

0.127	-0.929	1.35
-1.23	0.315	1.19
0.291	-0.627	-1.05
0.383	-0.563	-2.92
-0.927	0.914	-0.140
-0.111	1.48	0.247
-0.423	1.26	0.315
-0.293	0.144	-1.17
0.670	1.88	0.732

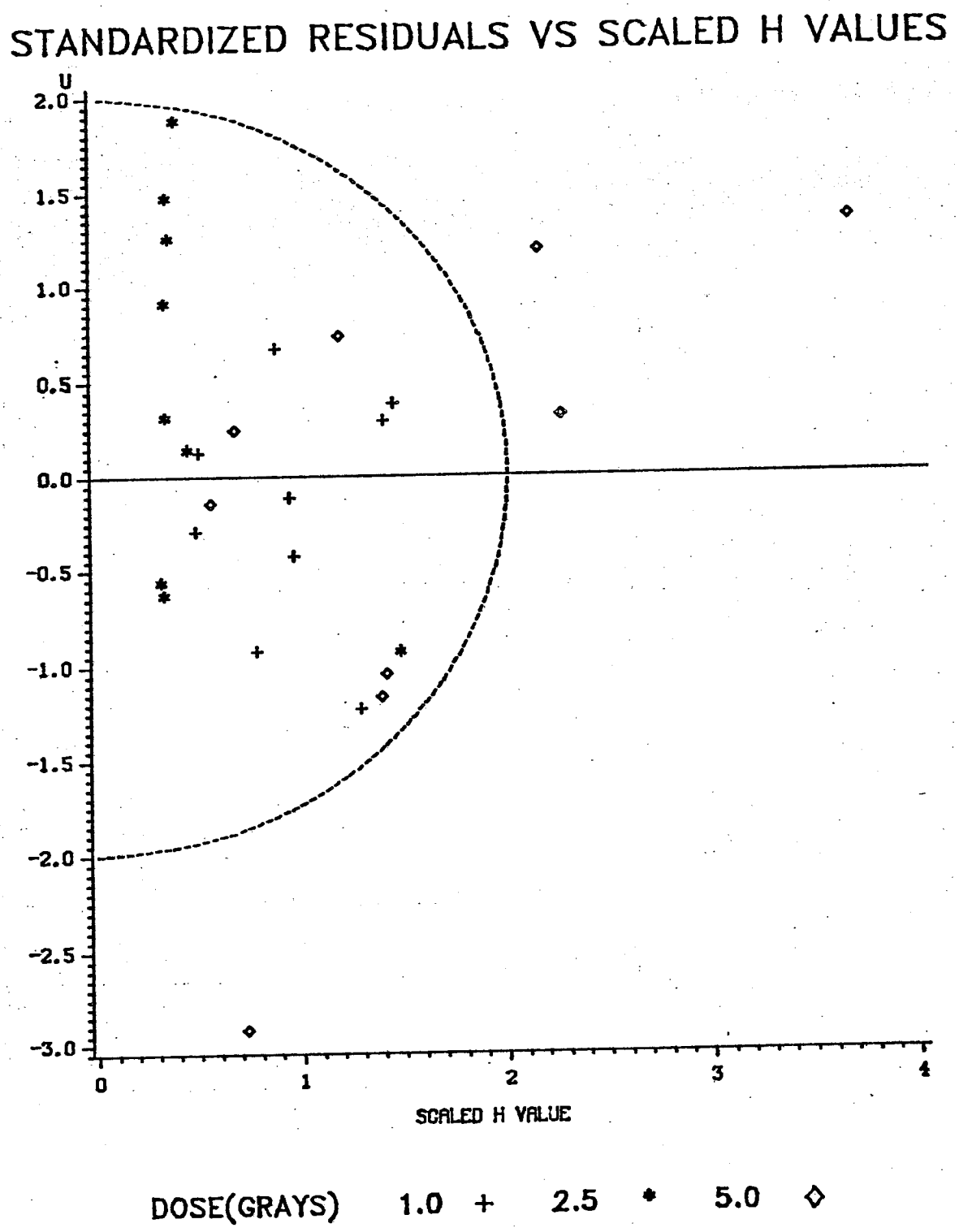

Figure 3. DIAGNOSTIC PLOT FOR DRA MODEL FOR CYTOGENETIC.
(See Section 2.3 and Table 10)

(b) Diagonal terms from the H matrix (p/n=0.111)

0.056	0.164	0.406
0.143	0.038	0.239
0.155	0.036	0.157
0.161	0.035	0.080
0.086	0.037	0.062
0.105	0.038	0.075
0.107	0.039	0.251
0.054	0.049	0.154
0.097	0.043	0.132

Note: Rows are dose rate and columns are dose groups (see Table 3)

Example 3: Lung Cancer Mortality (continued).

The product model (2.2) was used to obtain a convenient description of the association between lung cancer death rates and cigarette smoking. When the product model is adequate (i.e. the deviance is not large when compared to the appropriate chi-square distribution) than the "age-adjusted smoking effects" are equivalent to standardized risk ratios (with non-smokers as the referrent group). In some situations an additive model $\lambda_{jk} = \lambda_j + \phi_k$ ($\phi_1 = 0$) may be more appropriate. Note that the N *by* p "model matrix" X with rows X_i is identical to that for the product model and Poisson regression can be used to fit either of these GLFs. Both the product model and the additive model can be viewed as special cases of the GLF defined by equation (2.1). GLFs play an important role especially in exploratory statistical data analysis. In some situations, however, models have been proposed that cannot be conveniently expressed as a GLF. One such nonlinear function that has been proposed (see Whitemore and Keller, 1978) to describe the effect of dose rate (cigarettes/day) on age-specific lung cancer death rates is

$$\lambda(t,d) = (\gamma + \alpha d_k^{\theta})\, t_j^{\beta}\,, \tag{2.7}$$

where t is duration of smoking and d is exposure rate. Frome (1983) has provided a detailed discussion of this model which is intrinsically nonlinear in the parameters.

For the lung cancer data in Table 6 we define $t_j = (age-20)/42.5$ where age is the midpoint of the jth age interval, and d_k is the midpoint of the kth exposure (and we use $d_6=45$ for the last exposure group). The parameter γ represents the lung cancer death rate (per 10^5 man-years) in non-smokers ($d=0$) at age = 62.5 $t=1$), and γt^{β} corresponds to the age specific death rate in non-smokers at age t. A plot of the age specific death rates on a log-log scale will result in a straight line with a slope of β and intercept γ.

Equation (2.7) is a Weibull hazard function with one parameter β independent of d and t and the other parameter a function of dose rate: $\gamma + \alpha d^{\theta}$. Doll (1971) suggested that the hazard rate is approximately proportional to d and to the fourth power of duration of smoking (i.e. $\theta=1$ and $\beta = 4$). Note that if $\theta \neq 1$, then the exposure-effect relation will be concave ($\theta<1$) or convex $\theta>1$) toward the exposure axis. For estimation purposes we use

$$\lambda(X_i,\beta) = \left[\exp(\beta_2 + \beta_3\, x_{i1}) + \exp(\beta_4)\right] \exp(\beta_1 x_{i1}), \tag{2.8}$$

where $X_i = (logt_i, logd_i)$ and $\beta = (\beta, \log\alpha, \theta, \log\gamma)'$. The IRLS procedure (see Appendix) is used to obtaine ML estimates for Kahn's data (in Table 6) and the results are summarized in Table 11. In another study (of cigarette smoking in British physicians) Doll and Hill obtained data similar to Kahn's data in Table 6 – see Frome (1983, Table 1). The ML estimates for both of these data sets are given Table 11. The deviance for Doll and Hill's data in 59.6 with 59 df

indicating a good fit, while the deviance for Kahn's data, 43.5 with 26 df suggests a considerably poor fit. Of particular concern with the data from Kahn is the estimate of θ which is less than one, indicating a concave dose-effect relation.

Table 11. Maximum Likelihood Estimates for Parameters Specified by a Nonlinear Model* for Lung Cancer and Cigarette Smoking Data

	Data source	
Parameter	Kahn	Doll and Hill
β	3.38(0.18)‡	4.46(0.33)
$\log \alpha$	2.62(0.21)	1.82(0.66)
θ	0.83(0.06)	1.29(0.20)
$\log \gamma$	2.61(0.13)	2.94(0.58)
Deviance	43.5	59.6
d.f.	26	59

* Death rate = $\lambda_{jk} = (\gamma+\alpha d_k^{\theta})t_j^{\beta}$,
where t = (age -20)/42.5, d = cigarettes per day

† Data from Kahn (14) and Doll and Hill (15,16)

‡ Standard deviation in parentheses

3. BINOMIAL REGRESSION

In this section two examples will be presented to illustrate various aspects of binomial regression analysis. Example 4 provides an example of a linear regression function with replication at each set of experimental conditions. In this example the ANOVA-like table for binomial data will be used to test for "lack-of-fit" of the regression function and the assumption binomial variation. Example 5 will be used to illustrate the use of nonlinear functions (probit, logistic, and Weibull), and regression diagnostics for binomial data.

3.1 Linear Regression with Parallel Counts

Example 4: Streptonigrin Dose-Response Curve The data in Table 12 were obtained by DuFrain *et al* (1982) as part of a study that was undertaken to investigate the potential toxicity of a chemical clastogen (streptonigrin) on somatic cells and germ cells from female rabbits. The dependent variable y_{jk} is the number of damaged cells (lymphoblasts) for the *jth* animal exposed to streptonigrin dose d_k. One hundred cells were examined for each animal and we let $c_{jk} = 1$ (i.e. unit = 100 cells) for each observational unit, so that $\bar{y}$ is in per cent. The data are shown graphically in Figure 4. We assume that the y_{jk}s are independent and follow the binomial distribution with expectation $\mu_{jk} = c_{jk}\ \lambda(X_i,\beta)$ where

$$\lambda(X_i,\beta) = \beta_1 + \beta_2 d_j \tag{3.1}$$

This is a special case of the GLF

$$\lambda(X_i,\beta) = X_i\beta = \sum_j \beta_j x_{ij}, \tag{3.2}$$

with $x_{i1} = 1$ and $x_{i2} = d_j$ if the animal is in dose group j. Equation (3.1) is a linear regression function with intercept β_1 and slope β_2, and is used to describe these data for

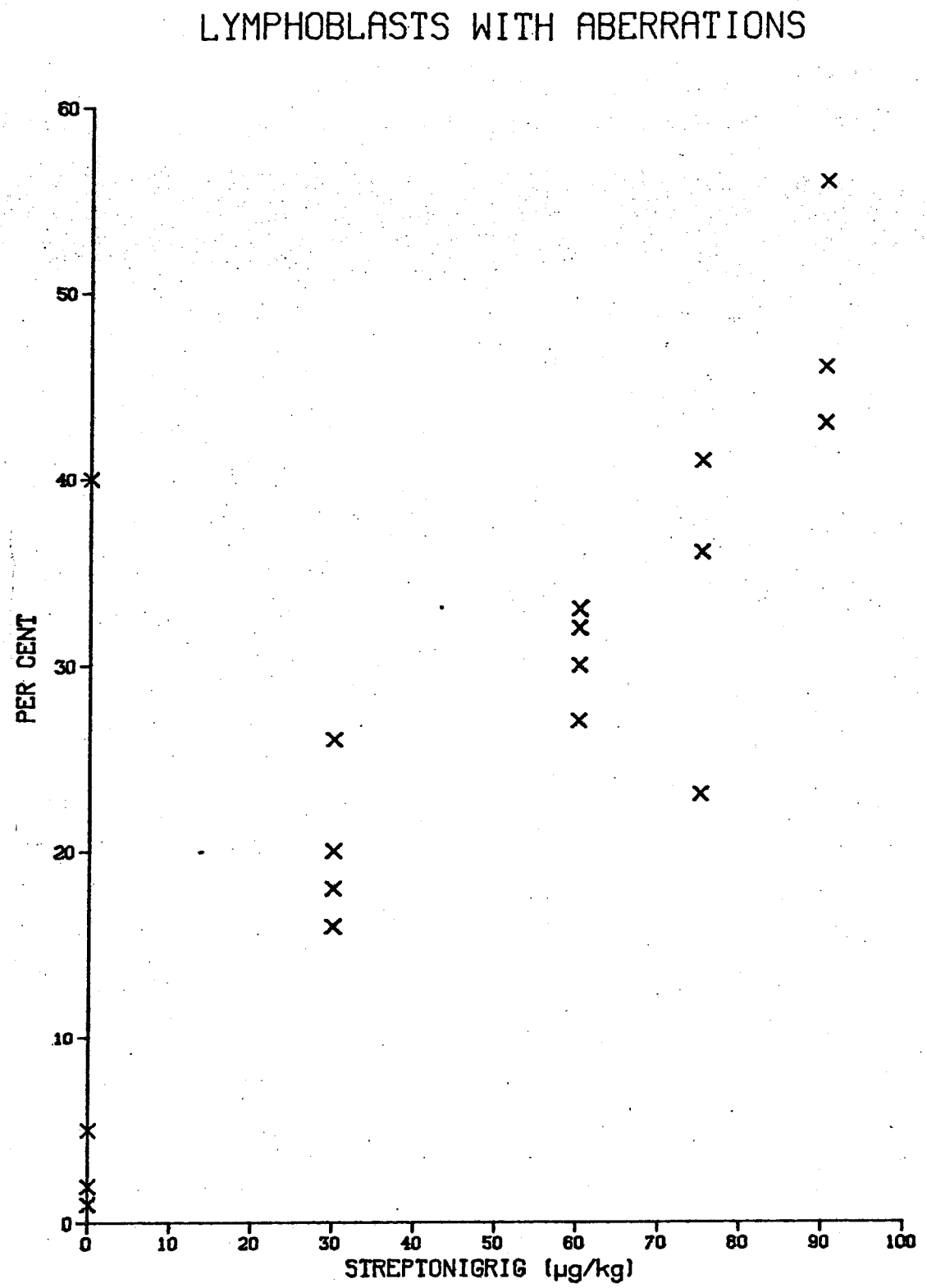

Figure 4. STREPTONIGRIN DOSE RESPONSE DATA.
(See Section 3.1 and Table 12)

$d < 100$ $\mu g/kg$. The ML estimates are obtained using IRLS (see the Appendix) and are $\beta_1^* = 2.67$ percent and $\beta_2^* = 0.478$ per cent per $\mu g/kg$, and their estimated standard deviations are 0.58 and 0.0189, respectively. The goodness-of-fit of this binomial regression model can be evaluated using the binomial ANOVA for these data (see Table 13a). Following the procedure described in the Appendix (see Table A.2) we first observe that deviance for the linear regression function (see Table 13a line 2) is 28.67 with 21 df (p = .12). This implies that neither the linear regression nor the assumption of binomial variation should be rejected. The assumption of binomial variation is further confirmed by the "within dose groups" deviance value of 22.64 with 18 df. The likelihood ratio statistic for "lack-of-fit" of the regression function is 6.03 with 3 df (p = .11) which again indicates that the linear regression function provides a reasonable model for these data. A test of the hypothesis $\beta_2 = 0$ can be based on the likelihood ratio statistic obtained as $D(\beta_0) - D(\beta_0 + \beta_1 d) = 336.6$ with 1 df ($p < 10^{-4}$).

Table 12. Streptonigrin Dose-Response Data*

Dose of Streptonigrin (μg/kg) d_j	Number of Rabbits n_j	Number of Lymphoblasts With Aberrations For Each Rabbit † y_{jk}					
0	6	5	2	1	1	2	4
30	5	20	16	26	18	16	
60	6	27	32	32	33	30	33
75	3	41	23	36			
90	3	43	46	56			

* Data obtained from R. J. DuFrain, Medical and Health Science Division, Oak Ridge Associated Universities (see DuFrain *et al*, 1982, Table 1)

† 100 Lymphoblasts were scored for each animal

For illustrative purpose consider the "linear logistic" regression function, i.e.

$$\lambda(d_i) = \exp(\beta_1 + \beta_2 d_i)/[1 + \exp(\beta_1 + \beta_2 d_i)]. \tag{3.3}$$

This can be viewed as a special case of the GLF

$$\lambda(X_i,\beta) = \exp(X_i\beta)/[1 + \exp(X_i\beta)], \tag{3.4}$$

where $X_i = (1,d_i)$. The binomial ANOVA using logistic GLF for the streptonigrin data is given in Table 13b. The values of the deviance on lines 1 and 3 are, of course, the same in both tables. The deviance for the linear logistic function (3.3) is 56.46 with 21 df ($p<10^{-4}$) indicating "lack-of-fit" of the *binomial regression model.* This could be due to heterogeneity of variance (i.e. overdispersion relative to the assumed binomial distribution within dose groups) or "lack-of-fit" of the regression function (3.3). As we noted earlier the assumption of binomial variation cannot be rejected. Consequently we can further assess the lack of fit of (3.3) by comparing the likelihood ratio statistic $D_3 - D_2 = 33.82$ with the chi square distribution with 3 df. Clearly this provides strong evidence against the linear logistic dose-response function for these data.

In summary, the streptonigrin dose-response data provides a simple example of binomial regression. A binomial regression model is based on two assumptions:

i) the y_{jk} are independent and follow the binomial distribution with expected value $\mu_{jk} = c_{jk}\lambda(X_j,\beta)$

ii) the regression function is specified.

If repeated observations are obtained at each set of experimental conditions then the appropriateness of the assumption of binomial variation can be tested. In this example we could have based our analysis on the totals at each dose -- see Table 14. The values of deviance in Table 14b differ from those in Table 13a by a constant amount, and in studies that involve a large number of experimental conditions this result can be used to simplify the analysis. In situations where heterogenity of variance is detected a detailed analysis (using standardized residuals within dose groups) should be considered. This could lead to the identification of spurious data values or the identification of additional factors that influence the dependent variable. When unexplained overdispersion occurs the analysis can be modified by introducing a heterogeneity factor (see the Appendix).

Table 13. Binomial ANOVA for Streptonigrin Data in Table 12

a) Linear Regression

	Regression Function †	df	Deviance	Likelihood Ratio Statistic	df
1.	β_1	22	365.3		
2.	$\beta_1+\beta_2 d$	21	28.67		
				6.03	3
3.	Dose Groups	18	22.64		
				22.64	18
4.	Complete	0	0.0		

b) Linear Logistic Regression

	Regression Function †	df	Deviance	Likelihood Ratio Statistic	df
1.	β_1	22	365.3		
2.	$\beta_1+\beta_2 d$	21	56.46		
				33.82	3
3.	Dose Groups	18	22.64		
				22.64	18
4.	Complete	0	0.0		

* The deviance $D(\beta^*)$ is used as a measure of residual (unexplained) variation (see Appendix eq. A.15).

† See explanation in text

Table 14. Streptonigrin Dose-Response Data

a) Totals from Table 12

Dose d_i	Number of Cells c_i	Number with Aberrations y_i	$\bar{y}_i$ (per cent)	$\lambda^*(d)$*
0	600	15	2.5	2.7
30	500	96	19.2	17.0
60	600	187	31.2	31.3
75	300	100	33.3	38.5
90	300	145	48.3	45.7

*$\lambda^*(d) = 2.67 + 0.478d$

b) Binomial ANOVA For Totals in Table 14a

Regression Function	df	Deviance	Likelihood Ratio Statistic	df
β_1	5	342.6		
			336.6	1
$\beta_1+\beta_2 d$	4	6.03		
			6.03	3
complete	0	0.0		

3.2 Nonlinear Regression Functions

Example 5: Chronic Bioassay of 2-AAF The data in Table 15 are from a study at the National Center for Toxicological Research that was undertaken to investigate the effect of a chemical compound 2-Acetylaminofluorene (2-AAF) on carcinogenesis (see Farmer, *et al*, 1979). Mice were continuously fed 2-AAF *ad libitum* in the diet at various concentrations from weanlings until they were either sacrificed, became moribund or died. Let y_{jk} denote the number of mice with liver neoplasms, and c_{jk} the number of animals at risk in the *kth* dose (exposure rate) group and the *jth* time interval group. Then $\bar{y}_{jk}$ is the proportion of mice with liver neoplasm and the regression function $\lambda(X_i,\beta)$ will represent the probability of observing a liver neoplasm in the *ith* group (where $i=8(j-1)+k$). For illustrative purposes we will begin our analysis using a linear-logistic regression function (3.4), and the factors DG ("dose" group with 8 levels) and TI (time interval with 9 levels) will be used to define the covariate vector X_i (see Example 3). This example will also illustrate how the computer program GLIM (Baker and Nelder, 1978) can be used to carry out the necessary data manipulations and computations. Figure 5 shows a listing of a GLIM-3 program that was used to (i) define the levels of the factors TI and DG; (ii) define the covariate vectors D,T, XD, and XT (that will be discussed later); and (iii) read the y_{jk} and c_{jk} into the vectors Y and N, respectively. In GLIM terminology the assumptions of binomial variation and a logistic regression function (3.4) are specified as follows:

$ERR B N $LINK G.

The values of deviance in the binomial-ANOVA table (see Table 16) were obtained using the FIT directive

$$\$FIT : DG : TI : TI + DG.$$

The diagonal terms from the H matrix (A.16) are obtained for the linear logistic function as follows:

$$\$EXTR\ \%VL\ \$CA\ H=\%VL * \%FV * (N-\%FV)/N.$$

The signed deviance residuals (see Appendix) are computed using the macro DVR (see Figure 5), and Figure 6 is a diagnostic plot of these standardized residuals and the scaled h_i for the linear logistic model TI + DG. The binomial ANOVA in Table 16 shows that most of the variation (96%) in these data can be explained by the linear logistic function (with TI and DG as factors), but the biologic interpretation of this model is not apparent.

Another approach that can be used to describe the data in Table 15 is to use (in GLIM terminology) a "probit link function". That is, we assume that $\lambda(X_i,\beta) = \Phi(X_i\beta)$, where Φ is the standard normal integral. The ANOVA table for the probit model is shown in table 17. Comparing the deviance values for the two factor models (TI + DG) suggests that the logit link provides a better representation that the probit link. We have not, however, used the values of the dose and time variables associated with the levels of the factors. Consider for example the regression function

$$\lambda_{jk} = \Phi[\delta_k + \beta\log(t_j)], \tag{3.5}$$

where t_j = time on study (in years). This corresponds to a log-normal response time distribution for each dose group, with constant standard deviation (β^{-1}) and mean $\mu_k = -\delta_k/\beta$. The deviance for (3.5) is 133.1 with 63 df (see DG + XT in Table 17). Farmer *et al* (1979) used the probit model in their analysis of these data and went further by assuming that

$$\lambda_{jk} = \Phi[\alpha+\theta\log d_k+\beta\log t_j] \quad . \tag{3.6}$$

They limited their analysis using (3.6) to the last three time points (18, 24, and 33 months) and eliminated the control ($d_1=0$) group. Farmer *et al* (1979) concluded from their limited analysis that (3.6) provided a "good fit" but indicated that the meaning of the slopes in (3.6) was not clear. Clearly, the model (3.6) cannot be considered as a reasonable representation of these data since it implies a zero response probability for unexposed animals and 45 of the mice in the zero dose group developed liver neoplasms. It would, of course, be possible to define other covariates for the probit (or the logit) function.

Another approach is to note that in both situations we have selected a cumulative distribution function (CDF) for the function G ( ), which insures that $0 \leq \lambda(X_i,\beta) \leq 1$. Another distribution function that has some appeal in this situation is the Weibull distribution, i.e.

$$\lambda_{jk} = 1 - \exp[-\ \delta_k t_j^{\beta}\ ]. \tag{3.7}$$

This model implies that "time to tumor" follows a Weibull distribution with scale parameter δ_k and shape parameter β that is independent of dose. The tumor incidence rate for each dose group is $\beta\delta_k t^{\beta-1}$, i.e. the exposure to 2-AAP (at constant rate) has a multiplicative effect on the tumor incidence rate. This relation is predicted, for example, by the multistage theory of carcinogenesis (see Whittemore and Keller, 1978). For estimation purposes the Weibull regression function (3.7) can be written as

$$\lambda_{jk} = 1 - \exp\left[-e^{\alpha_k+\beta\log(t_j)}\right] \tag{3.8}$$

or more generally as

$$\lambda(X_i,\beta) = 1 - \exp[-e^{X\beta}\ ]. \tag{3.9}$$

Table 15. Liver Neoplasms in Dead, Moribund and Sacrificed Mice Fed 2-AFF Continuously

Months on Study	DOSE (ppm) 0	30	35	45	60	75	100	150
9	0[1]/199[2]	1/147	1/76	0/52	0/345	0/186	1/168	1/169
12	0/164	1/51	2/27	1/14	2/283	0/153	3/149	2/152
14	1/133	1/42	0/25	2/14	1/243	0/124	1/127	1/127
15	0/115	1/75	1/35	0/20	3/203	1/109	5/99	1/100
16	1/205	2/66	2/61	3/304	6/287	7/193	2/100	7/110
17	0/153	4/69	5/443	6/302	8/230	9/166	3/85	1/82
18	6/555	34/2014	20/1102	15/550	13/411	17/382	19/213	24/211
24	20/762	164/2109	128/1361	98/888	118/758	118/587	76/297	126/314
33	17/100	135/445	72/200	42/103	30/67	37/75	22/31	9/11

[1] Number of Bladder Neoplasms

[2] Number of mice examined in this group

Source: Farmer *et al* (1979)

Table 16. Binomial ANOVA for 2-AAF Data in Table 15 Using a Linear-Logistic Regression Function

Linear Terms	Unexplained Variation*	df
minimal	2203.6	71
TI	595.1	63
DG	1965.6	64
TI+DG	85.7	56

Table 17. Binomial ANOVA FOR 2-AAF Data in Table 15 Using a Linear-Probit Regression function

Linear Terms	Unexplained Variation	df
minimal	2203.6	71
TI	595.1	63
DG	1965.6	64
DG+XT	133.1	63
TI+DG	109.1	56

```
$SUBFILE DATA !  10 FEB 83
$M TITLE  PBEX5: J.ENV.P&T(1979,P.57) ED-01 STUDY$E
$M VLIST  !   LIST OF VARIABLES
     Y= NUMBER OF LIVER NEOPLASMS  !
     N= NUMBER OF MICE!
     D= DOSE OF 2-AAF (PPM) CONTINUOUS FEEDING !
     T= MONTHS ON STUDY !
     XD= LOG( D ) !
     XT= LOG( T/12 ) -- LOG TIME IN YEARS!
     TI= TIME INTERVAL FACTOR ( ROWS OF TABLE )!
     DG= DOSE GROUP FACTOR ( COLUMNS OF TABLE )$E
$UNITS 72 $FAC TI 9 DG 8 !
$CA TI=%GL(9,8) : DG=%GL(8,1) : %R=9 !
$DATA 8 DOSE $READ 0.0 30 35 45 60 75 100 150
$CA D=DOSE(DG) : XD=%LOG(D) ! !
$DATA 9 TIME $READ 9 12 14 15 16 17 18 24 33
$CA T=TIME(TI) : XT=%LOG(T/12) !
$DATA Y $READ
 0  1 1 0 0 0 1 1
 0 1 2 1 2 0 3 2
 1  1 0 2 1 0 1 1
 0 1 1 0 3 1 5 1
 1 2 2 3 6 7 2 7
 0 4 5 6 8 9 3 1
 6  34 20 15 13 17 19 24
 20  164 128 98 118 118  76 126
 17 135 72 42 30 37 22 9
$DATA N $READ
 199  147 76 52 345 186 168 169
 164  51 27 14 283 153 149 152
 133  42 25 14 243 124 127 127
 115  75 35 20 203 109 99 100
 205  66 61 304 287 193 100 110
 153  69 443 302 230 166 85 82
 555   2014 1102 550 411 382 213 211
 762   2109 1361 888 758 587 297 314
 100   445 200 103 67 75 31 11
!     NOTE   N(33,35) CHANGED TO 200
!            N(12,30) CHANGED TO 51
$PR TITLE :: VLIST  $ERR B N $YVAR Y $DISP M !
$M DVR  $CA DV= -2*( Y*%LOG(%FV/Y) +!
   (N-Y)*%LOG( (N-%FV)/(N-Y) ) ) !
   $CA DV=%SQRT(DV) :DV=%IF(  %LT(Y,%FV),-DV,DV)$E
$RETURN
```

Figure 5. GLIM-3 PROGRAM FOR 2-AAF DOSE-TIME-RESPONSE DATA. (See Section 3.2 and Table 15)

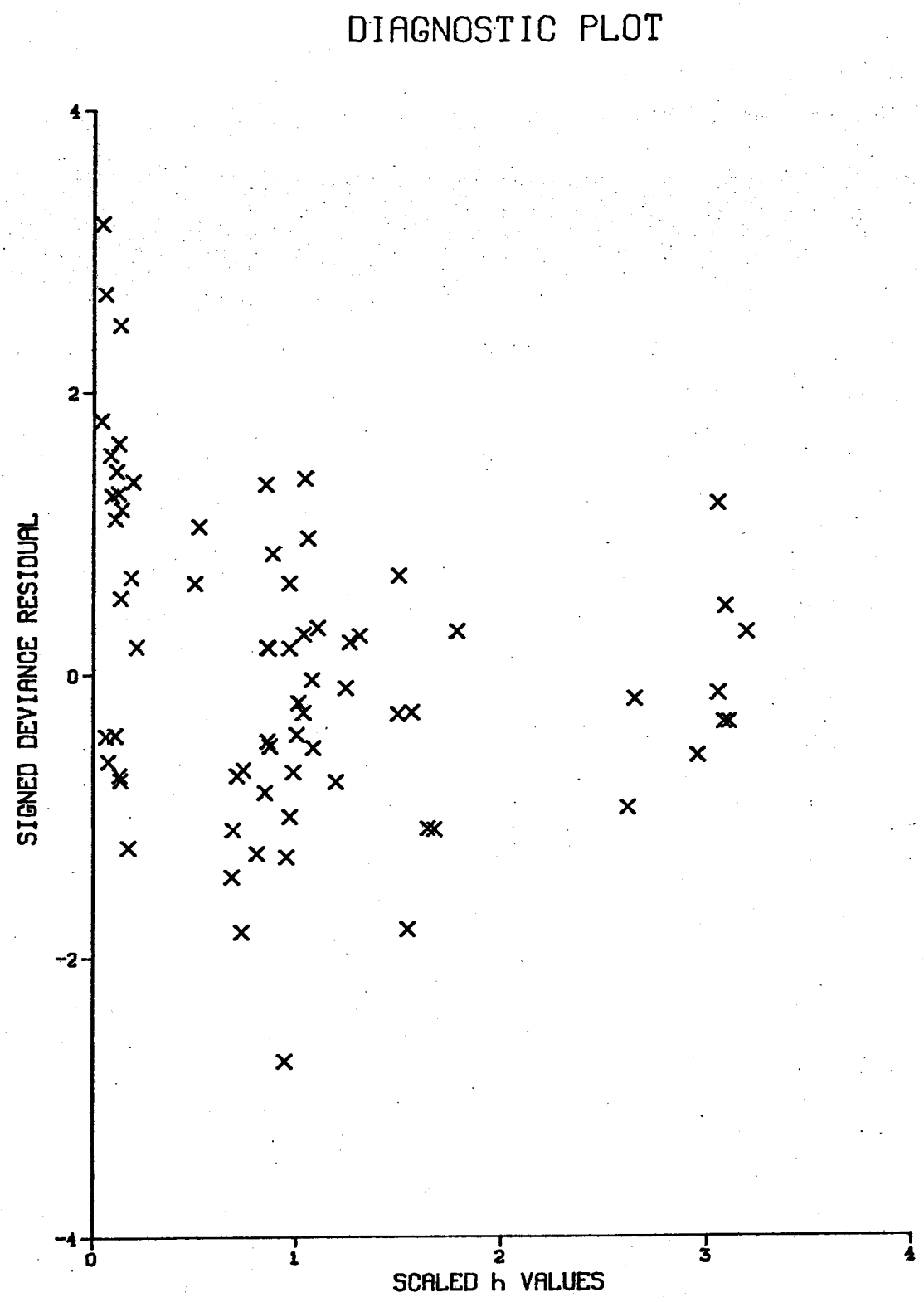

Figure 6. DIAGNOSTIC PLOT FOR 2-AAR DATA FOR LINEAR-LOGISTIC FUNCTION. (See Section 2.3 and Figure 5)

In GLIM terminology (3.9) can be used by specifying a complementary log-log link function, i.e. $LINK C. The ML estimates of the parameters for the Weibull model (3.8) are given in Table 18. The results of fitting various Weibull regression functions using GLIM are given in Table 19, where for example (3.8) is written as XT + DG, where DG is the factor dose group and XT = log (t_j) -- see Figure 5.

The deviance can be used as a measure of unexplained variation (similar to the residual sum of squares for normal linear models) for binomial data (but see Efron, 1976 for other possibilities). The primary purpose of Table 18 is to provide a summary measure for each model that is considered. In some circumstance the deviance is asymptotically distributed as a chi-square, but with small expected values this result is of limited value. As a rule of thumb, it is reasonable to assume that a model provides a "good fit" if the deviance is about equal to its df. The difference of the deviance for nested models provides a likelihood ratio statistic. For example in Table 18 the linear predictor on line 7 can be written as (in subscript notation)

$$\alpha + \beta \log t_j + \theta_1 d_k + \theta_2 d_k^2, \tag{3.10}$$

and line 6 correspond to (3.10) with $\theta_2 = 0$. The difference of the deviance for these two lines 129.1-98.1 = 31 with 1 df provides a test statistic for the $H_0{:}\theta_2 = 0$. The difference of the deviance values on lines 7 and 8 (98.1 - 96.7 = 1.4 with 5 df) is a test statistic for the constraint

$$\alpha_k = \alpha + \theta_1 dk + \theta_2 d_k^2,$$

and indicated that the α_k can be represented by a second degree polynomial in dose.

In summary, the 2-AAF dose-time-response data in Table 15 provides an example of a complex situations in which binomial regression analysis can be used. The analyst is confronted with the task of selecting a reasonable regression function $\lambda(X_i,\beta)$ to describe the effect of the administered compound (2-AAF) on a specific carcinogenic endpoint (liver neoplasms). This example clearly demonstrates that many different regression functions can be used to describe these data. Even with the additional constraint that our objective is to produce a parsimonious models, we still are unable to establish a "best" model on statistical grounds. This exploratory analysis does however strongly suggest that the "linear-probit" regression function is not appropriate for these data. Both the logistic and the Weibull CDF provide feasible GLFs for this example, and further analysis should include an attempt to establish some biologic grounds for either of these "time to tumor" distributions for this situation. Further, we should not lose site of the practical goals of this study, i.e. to investigate the relationship between length and level of dosing on the development of carcinogenic endpoints, and how this would affect regulatory decisions.

Table 18. ML Estimate For the Weibull Model (3.8) for the Data in Table 15

	α_1	α_2	α_3	α_4	α_5	α_6	α_7	α_8	β
Parameter									
Estimate	-6.75	-5.84	-5.65	-5.47	-5.18	-4.93	-4.53	-4.01	4.78
St. Dev.	.17	.11	.11	.12	.11	.11	.12	.11	.12

Table 19. Binomial ANOVA for 2-AAF Data in Table 15 Using Linear-Weibull Regression Function

	Linear Term	p	Unexplained Variation*	df
1	minimal	1	2203.6	71
2	TI	9	595.1	63
3	DG	8	1965.6	64
4	TI+DG	16	85.6	56
5	XT	2	615.2	70
6	XT+D	3	129.1	69
7	XT+D+D2	4	98.1	68
8	XT+DG	9	96.7	63
9	XT+D+D2.TI	12	77.31	60
10	TI+D+D2.TI	19	68.67	53

* Deviance is used as a measure of unexplained variation (see Appendix).

4. SUMMARY

In this paper we have shown how the LS principle can be used as a conceptual basis for fitting regression functions to discrete data that follow the Poisson or binomial distribution. The generalized LS estimates are obtained by solving a p dimensional system of equations (1.3) using an IRLS procedures. When the weights in the IRLS procedure are based on the Poisson or binomial assumption, the IRLS algorithm will yield a root of the likelihood equations, i.e. the generalized LS estimates are also ML estimates (see the Appendix). Consequently, Poisson regression and binomial regression (i.e. ML estimation under the Poisson or binomial assumption) can be viewed as part of the regression analysis paradigm. This considerably broadens the scope of regression analysis as a "methodologic paradigm", i.e. a scientific achievement which attracts

adherents from other disciplines (see Dolby, 1982). The analyst is therefore challenged to appeal to his general knowledge to develop "conjecture-based" models with data available for possible refutation. For example, the test for "lack of fit" of a regression function provides a probabilistic basis for evaluating the falsifiability of a proposed model. Dolby (1982) (in discussing the views of Karl Popper and Thomas Kuhn on the methodology of science) emphasizes the importance of "global conjectures", that are the province of the researcher in a particular field, as a basis for establishing specific (local) statistical hypothesis for attempted falsification. The alternative is the "exploratory investigation" of a specific data set that leads to an analysis that is descriptive rather than inferential. Finch (1979) points out that in some circumstances the extrapolation of a good description of the data that we have is the best we can do. The role of exploratory analysis is one of hypothesis generation, i.e. the preliminary investigation of data to uncover "good" descriptions that are relevant to the context. Attempts to test the goodness of fit of a model that is obtained in this way is misleading, and should only be used to provide guidance in an analysis. The five examples presented in this paper were selected to illustrate how regression analysis can be used for both types of data analysis, i.e. those based on global conjectures and exploratory techniques. In both situations the application of statistics to the life sciences is bound to be most fruitful when the analysis is based on a collaborative effort.

The general Poisson regression models include linear, log-linear, and intrinsically nonlinear regression functions. A numerical example from cytogenetic dosimetry was used to illustrate multiple linear regression for Poisson data. A more general dose-response model derived from the theory of dual radiation action was (DRA) also considered. The DRA regression function is intrinsically nonlinear in the parameters. Another important area where Poisson regression models are used is in the analysis of rates from observational studies. An example from an epidemiologic follow up study with the data organized into a life-table type of format was presented, and preliminary analysis was based on log-linear models. A nonlinear model, derived from the multistage theory of carcinogenesis, was then used to analyze lung cancer death rates among individuals who were regular cigarette smokers.

Binomial regression models are used for the analysis of binary (or quantal) response data, i.e. for situations where the outcome is one of two possible values (e.g. success or failure). Two numerical examples were presented that illustrate various aspects of binomial regression analysis. In the first example a linear dose response curve was used to describe the effect of streptonigrin on rabbit lymphoblasts. A linear regression function was used to describe the dose-response curve, and procedures for testing the assumption of binomial variation and lack-of-fit of the regression function were illustrated. In the second example mice were continuously fed a carcinogen (2-Acetylaminofluorene) for an extended period of time. Groups of mice were examined at various time points for each of several exposure levels and the number of mice with liver neoplasms was determined. These data were used to illustrate the application of several well-known regression functions (logistic and probit) to binomial data.

These examples of Poisson and binomial regression analysis are presented to illustrate situations in the biomedical sciences where discrete data that may follow either the Poisson or binomial distribution are encountered. The important problem in any specific situation is to determine an appropriate regression function that describes the effect of one or more covariates on the response. Historically, regression functions have been of the generalized linear type, a choice that appears to be based primarily on computational convenience. The computational requirements for the more general regression models are sufficiently complex that, in most situations, a computer based analysis is required. High quality, inexpensive portable programs (such as GLIM-3) are now widely available and can be used for all of the analyses described in this paper. The IRLS procedure (described in the Appendix) can be easily coded in any of the higher level languages (e.g. FORTRAN, Pascal) that are widely available on micro (personal) computers. Consequently, computational complexities should no longer limit the usefulness of Poisson and binomial regression models in routine data analysis.

APPENDIX

Equivalence of ML and IRLS for Poisson and Binomial Regression

The purpose of this Appendix is to show the equivalence of ML and IRLS for Poisson and binomial data. Charnes, Frome, and Yu (1976) have demonstrated the equivalence of ML and IRLS for situations where the dependent variable is from a member of the regular exponential family, and the regression function is in general nonlinear in the parameters. In the discussion that follows we will limit our discussion to the Poisson and binomial distribution.

Poisson and Binomial Regression Models

Let $y_1, y_2, \ldots, y_N$ denote the observed values of a random sample of size N from a population with density $h(y_i;\mu_i)$, where $\mu_i = \mu(X_i,\beta)$ denotes the expected value of Y_i. For the Poisson distribution

$$h(y;\mu) = e^{-\mu}\ \mu^y/y!\ , \qquad y=0,1,\ldots$$

and for the binomial distribution

$$h(y;\mu) = \binom{c}{y}(\mu/c)^y\ (1-\mu/c)^{c-y}, \qquad y = 0,1,\ldots,c.$$

The expected valueof Y_i is expressed as

$$\mu_i = \mu(X_i,\beta) = c_i\lambda(X_i,\beta),$$

where c_i is a known constant ("sample size") and $\lambda(X,\beta)$ is a known function, that we refer to as the regression function. The regression function describes the relation between the covariate vector $X_i = (x_{i1},\ldots,x_{im})$, and the unknown parameters $\beta = (\beta_1,\ldots,\beta_p)'$. Given the data $\{y_i,\ X_i,\ i=1,\ldots,N\}$ the problem is to obtain estimates of the parameters $\beta_1,\ldots,\beta_p$.

Maximum Likelihood Estimation

The logarithm of the likelihood function of β is

$$L(\beta) = \sum_i \log\ h(y_i;\mu_i).$$

Since Y is a random variable with a densiy function of the regular exponential family

$$h(y;\mu) = \exp\ \{\ yb(\mu)\ -\ q(\mu)\ +\ g(y)\ \}, \tag{A.1}$$

where $E(Y) = \mu$, and $b(\ )$ and $q(\ )$ are given in Table A-1 for the Poisson and binomial distribution. Following the approach of Charnes, Frome, and Yu (1976) differentiation (with respect to μ) on both sides of $\int h(y;\mu)dy = 1$ *yields*

$$E(Y) = q'(\mu)/b'(\mu) = \mu \tag{A.2}$$

where $b'(\mu)$ and $q'(\mu)$ denote derivatives with respect to μ. A second differentiation of the integral, along with evaluation of the derivative of (A.2) results in

$$V(Y) = b'(\mu)^{-1}. \tag{A.3}$$

These results are summarized in Table A-1.

The kernel of the log-likelihood function can be written

$$L(\beta) = \sum_i\Big\{y_i b[\mu(X_i,\beta)]\ -\ q[\mu(X_i,\beta)]\Big\}. \tag{A.4}$$

The ML equations are

$$\partial L/\partial\beta_j = \sum_i b'[\mu(X_i,\beta)](\partial\mu(X_i,\beta)/\partial\beta_j)$$
$$-\ \sum_i\ q'\ [\mu(X_i,\beta)](\partial\mu(X_i,\beta)/\partial\beta_j) = 0,\ j=1,\ldots,p.$$

By using (A.2) and (A.3) we obtain

$$\partial L/\partial\beta_j = \sum_i [V(Y_i)^{-1}[y_i - \mu(X_i,\beta)](\partial\mu(X_i,\beta)/\partial\beta_j)] \tag{A.5}$$

$$j = 1,\ldots,p$$

The likelihood equations are nonlinear with respect to the unknown parameters and an iterative procedure can be used to obtain a root of (A.5). A convenient computational approach to this problem is obtained by using IRLS (see below).

Iterative Reweighted Least Squares

Let $\bar{y}_i = y_i/c_i$ and w_i denote a positive weight that is proportional to the reciprocal of the variance of $\bar{y}_i$. For the binomial distribution c_i is the sample size and $\bar{y}_i$ is a proportion, i.e. the proportion of successes in c_i trials. For the Poisson distribution $\bar{y}_i$ is a "rate", (e.g. number of failures per unit time, number of events per unit area, etc.) and c_i is this "size" of the sample (e.g. number of time units, number of unit areas, etc.). Consider the following weighted sum of squares

$$S(\beta) = \sum_i w_i \; [\; \bar{y}_i - \lambda(X_i,\beta) \;]^2. \tag{A.6}$$

The least squares principle can be used to obtain an estimate of β by solving the system of equations

$$\sum_i w_i \; [\; \bar{y}_i - \lambda(X_i,\beta)](\partial\lambda(X_i,\beta)/\partial\beta_j), \tag{A.7}$$

$$j = 1,\ldots,p$$

Since $\lambda(X_i,\beta)$ is in general nonlinear in the parameters an iterative procedure is required to obtain an estimate of β. On iteration $k+1$ we replace $\lambda(X_i,\beta)$ with the linear terms in a Taylor series expansion about the current estimate β^k

$$\lambda(X_i,\beta) = \lambda(X_i,\beta^k) + P_i^k \; \delta^k, \tag{A.8}$$

where P_i^k denotes the ith row of the $N \times p$ matrix of partial derivatives $p_{ij} = \partial\lambda(X_i,\beta)/\partial\beta_j$ evaluated at β^k, and $\delta^k = (\delta_1^k,\ldots,\delta_p^k)'$ is the "correction vector".

Using (A.8) in (A.7) and the appropriate weights

$$w_i = \frac{c_i}{\lambda(X_i,\beta)} \quad \textit{,Poisson weights,} \tag{A.9}$$

or

$$w_i = \frac{c_i}{\lambda(X_i,\beta)[1-\lambda(X_i,\beta)]} \quad \textit{,binomial weights,}$$

evaluate at β^k we obtain

$$\sum_i w_i \; [\; \bar{y}_i - \lambda(X_i,\beta^k) - P_i^k \; \delta^k]\Big[\partial\lambda(X_i,\beta)/\partial\beta_j\Big]_{\beta=\beta^k} \tag{A.10}$$

Equation (A.10) can be written as

$$\mathbf{A}(\beta^k)\delta^k = \mathbf{G}(\beta^k), \tag{A.11}$$

where $\mathbf{A} = \mathbf{P}' \; \mathbf{W} \; \mathbf{P}$, $\mathbf{G} = \mathbf{P}' \; \mathbf{W} \; \mathbf{Z}$,$\mathbf{W} = diag(w_i)$, and $\mathbf{Z} = \Big[\; \bar{\mathbf{y}} - \lambda(X_i,\beta)\Big]$, where all expressions that involve β are evaluated at the current estimate β^k.

The linear system of equations is solved for δ^k and the revised estimate $\beta^{k+1} = \beta^k + \delta^k$ is obtained. This IRLS procedure continues until some convergence criteria are satisfied.

The matrix A is the "information matrix" with elements

$$a_{js} = \sum_i (p_{ij}p_{is}w_i) \;,\; j,s = 1,\ldots,p.$$

The system of equations (A.10) obtained using the least squares approach is identical to that obtained using the ML principle (A.5) -- to see this note that $w_i = c_i/v(Y_i)$ and $\partial\mu(X_i,\beta)/\partial\beta_j = c_i\partial\lambda(X_i,\beta)/\partial\beta_j$. Consequently, if the IRLS procedure converges to a stable solution (convergence is not guaranteed) it will yield a critical point of the likelihood equations. The IRLS procedure just described is equivalent to using the method of scoring to find a root of the likelihood equations (A.5). Further conditions (see Charnes,Frome, and Yu, 1976) to assure that β^* is a global maximum of $L(\beta)$ are (i) $L(\beta)$ be pseudoconcave over the parameter space, and (ii) that β^* satisfies (A.5). This will occur if $L(\beta)$ is defined over a convex parameter space and both $b[\mu(X,\beta)]$ and $q[\mu(X,\beta)]$ are concave in β over the parameter space. It will be the unique global solution of at least one of the $y_i b[\mu(X_i,\beta)]$, $- q[\mu(X_i,\beta)]$, $i=1,\ldots,N$ is strictly concave over the parameter space.

Table A-1

Characteristics of Poisson and Binomial Regression Models

	Poisson	Binomial
$b(\mu)$	$\log\mu$	$\log[\mu/(c-\mu)]$
$q(\mu)$	μ	$-\log(1-\mu/c)$
$E(Y)$	μ	μ
$V(Y)$	μ	$\mu(1-\mu/c)$
Regression Function	$\lambda(X,\beta)$	$\lambda(X,\beta)$
(Interpretation)	(rate)	(probability)
Regression Weight (w)	$\frac{c}{\lambda(X,\beta)}$	$\frac{c}{\lambda(X,\beta)[1-\lambda(X,\beta)]}$
$\bar{y} = y/c$	observed rate	observed proportion

Note: c is the sample size for a given observational unit with covariate vector X. The regression function is the expected value of $\bar{y}$ and is used as the dependent variable in the IRLS procedure (see text).

Covariance Matrix for the ML Estimates

The large sample covariance matrix of the ML estimators is the inverse of the information matrix $A(\beta)$ -- equation (A.11). If β^* is a stable solution of the likelihood equations (A.5) then estimates of the elements of this matrix are obtained by replacing β by the ML estimate β^*. For GLFs $\lambda(X_i,\beta) = G(\eta_i)$ where $\eta_i = X_i\beta = \sum_j x_{ij}\beta_j$ and A can be written as

$$A(\beta) = X' \; V \; X,$$

where V is diagonal with $v_i = w_i(\partial G_i/\partial\eta_i)^2 \; G_i^{-1}$. It can be shown (see McCullagh, 1983) that if N^{-1} A (β) has a positive definite limit as $N\to\infty$, then

$$E(\beta^* - \beta) = O(N^{-1}) \tag{A.12}$$

and

$$N^{1/2} \; (\beta^* - \beta) \sim N_p(0,N\sigma^2 A(\beta)^{-1}) + O_p(N^{-1/2}) \tag{A.13}$$

The notation N_p denotes the p-variate Normal distribution and the remainder terms in (A.12) and (A.13) refers to the difference between the cumulative distributions of the statistic and the Normal. The dispersion parameter σ^2 is equal to 1 if the dependent variable follows the Poisson or binomial distribution.

Evaluating Goodness of Fit

The results of fitting a model to data can be viewed as replacing the y_i with a set of "explanations" the μ_i^* that are derived from the regression function $\lambda(X_i,\beta^*)$. A measure of the discrepancy between the y_i and the μ_i^* that is convenient for both Poisson and binomial data is the *deviance* (see Nelder and Wedderburn, 1972). The deviance component for the *ith* observation for the Poisson distribution is

$$d_i^2 = 2\left[y_i \log (y_i/\mu_i^*) - (y_i - \mu_i^*)\right], \tag{A.14}$$

and for the binomial distribution

$$d_i^2 = 2\left[y_i \log (y_i/\mu_i^*) + (c_i - y_i)\log[(c_i - y_i)/(c_i - \mu_i^*)]\right], \tag{A.15}$$

where $\mu_i^* = c_i\lambda(X_i,\beta^*)$. The deviance is then obtained by summing the individual components, i.e.

$$D(\mathbf{y},\mu^*) = \sum_i d_i^2 \quad .$$

When the y_i are assumed to follow the Normal distribution the deviance is

$$\sum_i (y_i - \mu_i^*)^2,$$

i.e. the "residual sum of squares". Consequently the deviance can be used to construct a table similar to that used in standard linear model theory and referred to as an ANOVA (analysis of variance) table. The simplest model of interest (minimal model) has one parameter μ. At the other extreme is the full model which has one parameter for each observation--i.e. $\mu_i^* = y_i$, and the deviance is zero. The minimal model is usually too simple and one is interested in this model for reference purposes, since it provides the maximum value of the deviance for a given set of data.

If $\lambda(X_i,\beta)$ is a specific regression function of interest with $\beta = (\beta_1, \ldots, \beta_p)'$, then the deviance for this model is obtained from the ML estimate β^*. The "goodness of fit" of the regression function is evaluated by comparing the observed value of the deviance with the χ^2 distribution. For Poisson or binomial data the deviance is distributed approximately as a χ^2 with $N-p$ *df* when the assumed regression function is appropriate.

The analysis of variance has been most widely used for Normally distributed data when two or more factors and their interactions are of interest. Extension of these methods to GLFs for dependent variables in the regular exponential family have been developed by Nelder and Wedderburn (1972). An ANOVA like table is constructed by fitting a sequence of models and recording the *df* and deviance for each model that is considered. For each model that is fitted to the data the difference of the deviance for that model and the previous model represents the variation accounted for by the new factor having eliminated those terms of above it in the table (see the examples in Section 2 and 3). Note that the relative importance of a specific factor depends on when it is entered into the model--this is the same problem that occurs for the classical linear model when non-orthogonality occurs. For GLFs we may fit an increasing sequence of models, say $\mu \in H_j$, $H_0 \subset H_1 \subset \ldots \subset H_j \ldots$, and the difference of the deviance has an asymptotic χ^2 distribution if the more restrictive hypothesis is true.

A special form of the ANOVA table is of interest in experimental studies that are used to investigate a specific dose response curve is shown in Table A2. We assume that the y_{jk} are independent and follow the Poisson or binomial distribution with

$$E(y_{jk}) = c_{jk}\ \lambda(X_j,\beta)\ ,\ j=1,\ldots,N,\ k=1,\ldots,\ n_j,$$

i.e. y_{jk} is the response for the *kth* "parallel count" (replication) for dose group *j*. The dose response curve is given by $\lambda(X_j,\beta)$, and the deviance on line 2 of Table A.2 will be distributed approximately as a chi-square with $\sum_j n_j - p$ *df* if the *regression model* (i.e. *both* the regression function and the Poisson or binomial assumption) is appropriate. If this test statistic is large it may be due to "lack of fit" of the regression function or heterogenity of variance (under

the Poisson or binomial assumption). In this situation D_3 may be compared with the chi-square distribution with $\sum_i n_i - N$ df. If this statistic is significantly large (indicating heterogenity of variance) then the ratio

$$\frac{(D_2-D_3)/(N-p)}{D_3/(\sum_j n_j-N)}$$

may be compared with the F distribution (approximate test). A significant value of this F statistic indicates lack of fit of the *regression function* $\lambda(X_i,\beta)$. We have partitioned the unexplained variation D_2 into two components since $D_2 = (D_2-D_3) + D_3$, where D_3 is equivalent to the "pure error" sum of squares in Normal regression analaysis (see Draper and Smith, 1966, chap. 2).

The values of the deviance can also be used to construct an R^2*—type* measure of variance explained, i.e.

$$R^2 = 100\ (D_1-D_2)/D_1$$

is the percent of the total variation (as measured by the deviance) that is explained by the regression function $\lambda(X,\beta)$.

Table A.2

ANOVA Table for Lack of Fit Test for Poisson or Binomial Data

	Regression Function	Number of Parameters	Deviance	df
1	minimal	1	$D_1 = D(y, \mu^* 1)$	$\sum_j n_j - 1$
2	$\lambda(X_i,\beta)$	p	$D_2 = D[y, \mu(X_i,\beta^*)]$	$\sum_j n_j - p$
3	Dose Groups	N	$D_3 = D[y, \mu(\lambda^*_j)^+]$	$\sum_j n_j - N$
4	Complete	$\sum_{i=1}^{N} n_i$	0.0	

$+\mu(\lambda_j^*)$ denotes a vector of fitted values based on the model $E(y_{jk}) = c_{jk}\ \lambda_j$. The ML estimates $\lambda_j^* = \sum_k y_{jk}/\sum_k c_{jk}$ are used to compute the deviance, i.e. $\mu_{jk}^* = c_{jk}\ \lambda_j^*$ on line 3.

Regression Diagnostics

An important area of regression analysis that has received considerable attention (for the standard linear model) in recent years is regression diagnostic -- see e.g. Belsey *et al* (198). Diagnostic procedures are used to check for outlying y-values and extreme points in the "model space". Extension of these techniques to binomial (logistic) regression models and Poisson regression models have also been proposed -- see Pregibon (19810, Frome (1983), and ( ). The basic "building blocks" that are required for various diagnostic measures are standardized residuals of some type and the diagonal terms, h_i, from the matrix

$$\mathbf{H} = \mathbf{W}^{1/2}\ \mathbf{P}\ (\mathbf{P}'\ \mathbf{W}\ \mathbf{P})^{-1}\ \mathbf{P}'\ \mathbf{W}^{1/2}, \qquad \text{(A.16)}$$

where all quantities that depend on β are evaluated at the ML estimate β^* (see equation A.11). The diagonal terms from this matrix are useful in detecting extreme points in the model space that may have a substantial influence on the fitted model. Recall that for the standard linear

model $H = X(X'X)^{-1}X'$, $I - H$ is the projection matrix, and large values of h_i identify extreme points in the model (design) space. For GLFs $\lambda(X_i,\beta) = G(\eta_i)$, where $\eta_i = \sum_j \beta_j x_{ij}$, and H can be written as

$$H = V^{1/2} X (X' V X)^{-1} X' V^{1/2},$$

where V is diagonal with $v_i = \{w_i(\partial G_i/\partial \eta_i)^2\}$. Note that $\sum_i h_i = p$ and that large values of h_i (say, greater than 2p/n) indicate extreme points in the model space. If u_i denotes a standardized residual the variance of u_i is approximately $1-h_i$ and "adjusted residuals" are given by $u_i/(1-h_i)^{1/2}$. Two possible definition of standardized residual are

$$u_i = (y_i - \mu_i^*)/var(y_i)^{1/2}, \text{ or}$$

$$u_i = sign(y_i - \mu_i) d_i,$$

where d_i^2 is deviance component (see A.14 and A.15).

Heterogeneity of Variance

In practical data analysis the assumption of binomial or Poisson variation may be unrealistic. Usually the variance will be greater than that predicted, a phenomenon referred to as overdispersion or heterogeneity of variance. In some situations it is reasonable to assume that $\mu_i = c_i\lambda(X_i,\beta)$ and that the variance of y is proportional to that predicted under the Poisson or binomial assumptions. In this situation the estimated parameter covariance matrix is multiplied by the dispersion parameter σ^2. An estimate of σ^2 is obtained as

$$\hat{\sigma}^2 = \frac{1}{N-p} \sum_i \frac{(y_i - \mu_i^*)^2}{var(y_i)},$$

where $var(y) = \mu^*$ for the Poisson distribution, and $var(y) = c\lambda^*(1-\lambda^*)$ for the binomial distribution. Estimates of the regression parameters can be obtained using maximum quasi-likelihood (MQL) estimation--see Wederburn (1974), McCullagh (1983). Apart from the multiplier σ^2 the quasi-likelihoods can be treated for the most part just like ordinary likelihoods. In particular the quasi-likelihood equations are given by (A.5) so that IRLS procedure can be used to obtain the MQL estimates for overdispersed data.

REFERENCES

BAKER, R. J. and NELDER, J. A. (1978), *Generalized Linear Interactive Modelling (GLIM), Release 3,* Numerical Algorithms Group, Oxford.

BELSLEY, D. A., KUH, E., and WELSCH, R.E. (1980), *Regression Diagnostics,* New York, Wiley.

BISHOP, Y. M. M., FINEBERG, S. E. and HOLLAND, P. W. (1975), *Discrete Multivariate Analysis: Theory and Practice,* Cambridge, Massachusetts: MIT Press.

BLISS, C. I. (1935), "The Calculation of The Dosage-Mortality Curve," *Annals of Applied Biology* 22, 134-67.

BRESLOW, N. E., LUBIN, J. H., MAREK, P. and LANGHOLZ, B. (1983) "Multiplicative Models for Cohort Analysis," *Journal of The American Statistical Association* 78, 1-12.

CHARNES, A., FROME, E. L., and YU, P. L. (1976), "The Equivalence of Generalized Least Squares and Maximum Likelihood Estimation in the Exponential Family," *Journal of the American Statistical Association,* 71, 169-172.

COCHRAN, W. G. (1940), "The Analysis of Variance When Experimental Errors Follow The Poisson or Binomial Laws," *Annals of Mathematical Statistics* 11, 335-347.

COCHRAN, W. G. (1983), *Planning and Analysis of Observational Studies,* New York: J. Wiley and Sons.

COX, D. R. (1970), *The Analysis of Binary Data,* London, Methuen.

DOLBY, G. R. (1982), "The Role of Statistics in the Methodology of Science," *Biometrics* 38, 1069-1083.

DRAPER, N.R. and SMITH, H., *Applied Regression Analysis,* New York: John Wiley & Sons, Inc., 1966.

DUFRAIN, R. J., LITTLEFIELD, L. G., JOINER, E. E. and FROME, E. L. (1980), "In Vitro Human Cytogenetic Dose Response Systems," in: *The Medical Basis for Radiation Accident Preparedness* (Hubner, K. F. and Fry, S. S. eds.), Elsevier North-Holland, New York.

DUFRAIN, R. J., LITTLEFIELD, L. G., WILMER, J. L., and FROME, E. L. (1982), "Evaluation of Chemically Induced Cytogenetic Lesions in Rabbit Oocytes, II," *Mutation Research,* 94, 103-114.

EFRON, B. (1978), "Regression and ANOVA with Zero-One Data: Measures of Residual Variation," *Journal of the American Statistical Association,* 73, 113-121.

FARMER, J. H., KODELL, R. L., GREENMAN, D. L. and SHAW, G. W. (1979), "Dose and Time Response Models For the Incidence of Bladder and Liver Neoplasms in Mice Fed 2-Acetylaminofluorene Continuously," *Journal of Environmental Pathology and Toxicology* 3, 55-68.

FINCH, P.D. (1979) "Description and Analogy in the Practice of Statistics," *Biometrika* 66, 195-208.

FROME, E. L. (1972), "Nonlinear Regression and Spectral Estimation in Biomedical Data Analysis," Ph.D. Dissertation, Emory University: Atlanta, Georgia.

FROME, E. L. (1981), "Poisson Regression Analysis," *American Statistician,* 35, 262-263.

FROME, E. L. (1982), "Fisher's Exact Variance Test for the Poisson Distribution," *Appl. Statistics,* 31, 67-71.

FROME, E. L. (1983), "The Analysis of Rates Using Poisson Regression Models," *Biometrics,* 39, 665-674.

FROME, E. L. and BEAUCHAMP, J. J. (1968), "Maximum Likelihood Estimation of Survival Curve Parameters," *Biometrics,* 24, 595-605.

FROME, E. L. and CHECKOWAY, H. (1984), "The Use of Poisson Regression Models in Estimating Incidence Rates and Ratios" *American Journal of Epidemiology* (to appear).

FROME, E. L. and DUFRAIN, R. J. (1983), "Maximum Likelihood Estimation for Cytogenetic Dose-Response Curves," Research Report 123, Oak Ridge National Laboratory, Oak Ridge, Tennessee.

FROME, E. L., KUTNER, M. H. and BEAUCHAMP, J. J. (1973). "Regression Analysis of Poisson-distributed Data," *Journal of the American Statistical Association,* 68, 935-940.

GART, J. J. (1978), "The Analysis of Ratios and Cross-Product Ratios of Poisson Variates With Application to Incidence Rates," *Communication in Statistics Theory and Methods,* A7 917-37.

GOODNIGHT, J. H. and SALL, J. P. (1982), "The NLIN Procedure," *SAS Users Guide: Statistics,* SAS Institute, Cary, North Carolina.

HAIGHT, F.A., (1967), *Handbook of the Poisson Distrubition,* New York: John Wiley & Sons, Inc.

HOLFORD, T. R. (1980), "The Analysis of Rates and Survivorship Using Log-Linear Models," *Biometrics* 36, 299-305.

JOHNSON, N. L. and KOTZ, S. (1970), *Distributions inStatistics:Discrete Distributions,* New York: J. Wiley and Sons.

KAHN, H. A. (1966) "The Dorn Study of Smoking and Mortality Among U.S. Veterans: Report on Eight and One-Hlf Years of Observation," in: *Epidemiologic Approaches to The Study of Cancer and Other Chronic Diseases,* ed. Haenszel, W., *National Cancer Institute Monograph* 19, 1-125.

KELLERER, A. M. and ROSSI, H. I. (1972), "The Theory of Dual Radiation Action," *Curr. Topics Radiat. Res.,* 8, 85-158.

KENDALL, M. G. and STUART, A., (1946) *The Advanced Theory of Statistics, Vol. 2,* New York: Hafner Publishing Company.

KOCH, G. G., ATKINSON, S. S. and STOKES, M.E. (1984), "Poisson Regression," in: *Encyclopedia of Statistical Sciences,* eds Kotz, S., Johnson, L. L., and Read, A, New York: J. Wiley & Sons (in press).

LEA, D. E. (1955), *Action of Radiations on Living Cells,* 2nd Edition, Cambridge Press.

McCULLAGH, P. (1983), "Quasi-Likelihood Functions," *The Annals f Statistics,* 11, 59-67.

McCULLAGH, P. and NELDER, J. A. (1983), *Generalized Linear Models,* London: Chapman and Hall.

MIETTIEN, O. S. (1972), "Standardization of Risk Ratios", *American Journal of Epidemiology* 96, 383-88.

MOORE, R. H. and ZEIGLER, R. K. (1967), "The Use of Nonlinear Regression Methods for Analyzing Sensitivity and Quantal Response Data," *Biometrics,* 23, 563-67.

NELDER, J. A. and WEDDERBURN, R. W. M. (1972), "Generalized Linear Models," *Journal of the Roy. Statistical Society, Series A 135,* 370-384.

PALMGREN, J. (1981), "The Fisher Information Matrix For Log-Linear Models Arguing Conditionally on the Observed Explanatory Variables," *Biometrika* 68, 563-66.

PREGIBON, D. (1981), "Logistic Regression Diagnostics," *The Annals of Statistics,* 9, 705-724.

PURROTT, R. J. and REEDER, E. (1976), "The Effect of Changes in Dose Rate on the Yield of Chromosome Aberrations in Human Lymphocytes Exposed to Gamma Radiation," *Mutation Research,* 35, 437-444.

RAO, C. R. (1965), *Linear Statistical Inference and Its Applications,* New York: John Wiley & Sons, Inc.

WHITLEMORE, A. and KELLER, J. B. (1978), "Quantitative Theories of Carcinogenesis", *SIAM Review 20,* 1-30.