PIRLS: Poisson Iteratively Reweighted Least Squares Computer Program for Additive, Multiplicative, Power, and Non-linear Models

Leif E. Peterson
Center for Cancer Control Research
Baylor College of Medicine
Houston, Texas 77030

November 3, 1997

Abstract

A computer program to estimate Poisson regression coefficients, standard errors, Pearson $\chi^2$ and deviance goodness of fit (and residuals), leverages, and Freeman-Tukey, standardized, and deletion residuals for additive, multiplicative, power, and non-linear models is described. Data used by the program must be stored and input from a disk file. The output file contains an optional list of the input data, observed and fitted count data (deaths or cases), coefficients, standard errors, relative risks and 95% confidence intervals, variance-covariance matrix, correlation matrix, goodness of fit statistics, and residuals for regression diagnostics.

Key Words: Poisson Regression, Iteratively Reweighted Least Squares, Regression Diagnostics, Matrix Manipulation.

1 INTRODUCTION

1.1 Poisson Distribution

Poisson regression models are commonly used for modeling risk factors and other covariates when the rates of disease are small, i.e., cancer [1, 2, 3, 4, 5]. An area of research in which Poisson regression of cancer rates is commonly employed is the investigation of radiation-induced cancer among Hiroshima and Nagasaki atomic bomb survivors [6, 7, 8] and individuals occupationally [9], medically [10] and environmentally [11] exposed to ionizing radiation. Poisson regression has also been used to investigate the effects of information bias on lifetime risks [12], diagnostic misclassification [13, 14], and dose-response curves for radiation-induced chromosome aberrations [15].

Several computer programs have been developed that perform Poisson regression [16, 17, 18, 19, 20]. While the PREG program [3] can fit additive, multiplicative, and non-linear models, it does not provide the capability to fit power models nor perform regression diagnostics. This report documents the PIRLS Poisson regression algorithm for fitting additive, multiplicative, power, and non-linear models and regression diagnostics using leverages of the Hat matrix, deletion and standardized residuals.

The critical data required to apply the computer code are:

- Number of cases for each stratum
- Number of person-years for each stratum
- Design matrix of independent variables

There is only one output format, listing the input data, observed and fitted count data (deaths or cases), coefficients, standard errors, relative risks (RR) and 95% confidence intervals, variance-covariance matrix, correlation matrix, goodness of fit statistics, and residuals for regression diagnostics. The program and all of its subroutines
are written in FORTRAN-77 and are run on a desktop PC with a AMD-K6-166MHz(MMX) chip. The algorithm
can be compiled and run on Unix machines as well. The PC version of the executable requires 180,224 bytes of
random access memory (RAM) to operate. Average execution time is several seconds per run, depending on the
amount of input data.

2 PROGRAM DESCRIPTION

The computer program proceeds as follows. During run-time, the user interface with the program involves the
following:

- Specifying the type of model to be fit (1-additive, 2-multiplicative, 3-power, and 4-non-linear)
- For additive models, specifying the power, ρ, which ranges from 1 for purely additive models to 0 for purely
  multiplicative models
- Specifying the number of independent variables (not cases or person-years) that are to be read from the
  input file
- Specifying the number of parameters to estimate
- For non-linear models, optional starting values for parameters
- For multiplicative models, specifying if the regression coefficients are to be exponentiated to obtain RR and
  its 95% CI
- Specifying whether or not the variance-covariance and correlation matrices are written to the output file
  whose name is specified by the user
- Specifying whether or not the input data are written to the output file whose name is specified by the user
- Specifying whether or not the observed and fitted count data (deaths or cases) are written to the output file
- Specifying whether or not the leverages, and standardized and deletion residuals are written to the output file

Once the data above have been specified during run time, the program does the following:

1. First, a “starting” regression is performed by regressing the ratio of counts (deaths or cases) to person-years
   of follow-up as the dependent variable on the independent variables. For non-linear models, the default
   starting regression can be skipped by specifying starting values for each parameter.

2. The difference between the observed and fitted counts (error residuals) from the starting regression are
   used as the dependent variable and regressed on the cross product of the inverse of the information matrix
   (variance-covariance-matrix) and the score vector to obtain the solution vector.

3. The solution vector is then added to the original regression coefficients obtained during the starting regres-
   sion.

4. Iterations are performed until the sum of the solution vector is below a fixed value called the convergence
criterion, for which the default in PIRLS is $10^{-5}$.

5. After convergence is reached, the standard errors are obtained from the diagonal of the variance-covariance
   matrix, and if specified, residuals for regression diagnostics are estimated.

6. Depending on specifications made by the user, various data are written to the output file, and then the
   program terminates run-time.

3 NOTATION AND THEORY

3.1 Binomial Basis of Poisson Regression

For large $n$ and small $p$, e.g., a cancer rate of say 50/100,000 population, binomial probabilities are approximated
by the Poisson distribution given by the form

$$ P(x) = \frac{e^{-m} m^x}{x!}, $$

(1)
where the irrational number $e$ is 2.71828.... In many respects, the Poisson distribution has been widely used in science and does not really have a direct relationship with the binomial distribution. As such, $np$ is replaced by $\mu$ in the relationship

$$f(x) = \frac{\mu^x e^{-\mu}}{x!},$$

where $\mu$ is the mean of the Poisson distribution. It has been said that as long as $npq$ is less than 5, the data are said to be Poisson distributed.

### 3.2 Poisson Assumption

The Poisson assumption states that, for stratified data, the number of deaths, $d$, in each cell approximates the values $x=0,1,2,...$ according to the formula

$$P(d = x) = \frac{(\lambda T)e^{-\lambda T}}{x!},$$

where $\lambda$ is the rate parameter which is equal to the number of deaths in each cell divided by the respective number of person-years ($T$) in that same cell. As an example, Table 1 lists a multiway contingency table for the British doctor data for coronary disease and smoking [21]. The rates for non-smokers, $\lambda_i(0)$, are simply the ratio of the number of deaths to $T_i$ in age group $i$ and exposure group 0, whereas the rates for smokers, $\lambda_i(1)$, are the ratio of deaths to $T_i$ in age group $i$ for the exposure group 1.

### 3.3 Ordinary Least Squares

Let us now look into the sum of squares for common linear regression models. We recall that most of our linear regression experience is with the model

$$Y = X\beta + \varepsilon,$$  

where $Y$ is a column vector of responses, $X$ is the data matrix, $\beta$ is the column vector of coefficients and $\varepsilon$ is the error. Another way of expressing the observed dependent variable, $y_i$, of (4) in model terms is

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} \cdots \beta_j x_{ij} + \varepsilon_i,$$

and in terms of the predicted values, $\hat{y}_i$, is

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} \cdots \hat{\beta}_j x_{ij} + e_i.$$  

The residual sum of squares of (6) is

$$SS(\beta) = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n e_i^2 = \varepsilon^T \varepsilon.$$  

### Table 1: Deaths from coronary disease among British male doctors [21].

| Age group, $i$ | $d_i$ | $T_i$ | $\lambda_i(0)$ | $d_i$ | $T_i$ | $\lambda_i(1)$ |
|---------------|------|------|---------------|------|------|---------------|
| 35-44         | 2    | 18,790 | 0.1064        | 32   | 52,407 | 0.6106        |
| 45-54         | 12   | 10,673 | 1.1243        | 104  | 43,248 | 2.4047        |
| 55-64         | 28   | 5,710  | 4.9037        | 206  | 28,612 | 7.1998        |
| 65-74         | 28   | 2,585  | 10.8317       | 186  | 12,663 | 14.6885       |
| 75-84         | 31   | 1,462  | 21.2038       | 102  | 5,317  | 19.1838       |
Because of the relationship
\[ e = Y - X\beta, \]  
we can rewrite (7) in matrix notation as
\[ SS(\beta) = (Y - X\beta)'(Y - X\beta). \]  

### 3.4 Partitioning Error Sum of Squares for Weighted Least Squares

Suppose we have the general least squares equation
\[ Y = X\beta + \varepsilon. \] 
and according to Draper and Smith [22] set
\[ E(\varepsilon) = 0, \quad V(\varepsilon) = V\sigma^2, \quad \varepsilon \sim N(0, V\sigma^2), \]
and say that
\[ P'P = PP = P^2 = PP = V. \]

Now if we let
\[ f = P^{-1}\varepsilon, \quad E(f) = 0, \]
and then premultiply (10) by \( P^{\dagger} \), as in
\[ P^{-1}Y = P^{-1}X\beta + P^{-1}\varepsilon, \]
or
\[ Z = Q\beta + f. \]

By applying basic least squares methods to (14) the residual sums of squares is
\[ f'f = \varepsilon'(P^{-1})'P^{-1}\varepsilon, \]

since
\[ (P^{-1}\varepsilon)' = \varepsilon'(P^{-1})', \]
and since
\[ (B^{-1}A^{-1}) = (AB)^{-1}, \]
then
\[ f'f = \varepsilon'(PP)^{-1}\varepsilon, \]
which is equivalent to
\[ f'f = \varepsilon'(V)^{-1}\varepsilon. \]

The sum of squares residual then becomes for (14)
\[ SS(\beta) = (Y - X\beta)'V(Y - X\beta). \]

### 3.5 Partitioning Error Sum of Squares for Iteratively Reweighted Least Squares

According to McCullagh and Nelder [23], a **generalized linear model** is composed of the following three parts:

- **Random** component, \( \lambda \), which is normally distributed with mean \( \frac{1}{\tau_i} \cdot \mu_i \) and variance \( \frac{1}{\tau_i^2} \).
- **Systematic** component, \( x \), which are covariates related to the linear predictor, \( (x_i'\beta) \), by the form
  \[ (x_i'\beta) = \sum_{j=1}^{p} x_{ij}\beta_j. \]
The link between the random and systematic components for a linear model is given by

$$\lambda_i = \lambda_i(x_i') = \lambda_i(x_i', \beta) = \exp(x_i')$$

where $\lambda_i$ is the observed rate parameter and $\lambda_i = (x_i', \beta)$ is the predicted rate parameter. It should be pointed out that the rate parameter, $\lambda_i = (x_i', \beta)$, serves as its own link function. Thus, the rate parameter provides a linearization of our distributional models so that we can use multiple linear regression to obtain Best Linear Unbiased Estimators (BLUEs). In effect, the rate parameter allows us to reduce non-linear (non-Gaussian) models to linear models so that we can solve the system of linear equations. The rate parameter can take on various linear and non-linear functions. These are shown in Table 2. The power transformation was adapted from [24].

As shown in Table 2, it is obvious that the random component, $\lambda_i$, can take on a non-linear relationship with the systematic component, $x_i$. To linearize $\lambda_i$ in a least squares approach by a succession of stages or iterations, we carry out a Taylor series expansion of $\lambda_i$ about a point $\beta_0$. Since we know that

$$g(\beta) = g(\beta_0) + g(\beta_0)'(\beta - \beta_0)$$

a Taylor series expansion of $\lambda_i$ about $(\beta - \beta_0)$ gives us

$$\lambda_i \simeq \lambda_i(x_i', \beta_0) + \sum_{j=1}^{p} \left[ \frac{\partial \lambda_i(x_i', \beta)}{\partial \beta_j} \right] \beta_j = \beta_0 (\beta - \beta_0) + \epsilon_i$$

Rearranging this equation, we have

$$\lambda_i - \lambda_i(x_i', \beta_0) = Z^0 \delta^0 + \epsilon_i$$

where

$$Z_{ij}^0 = \left[ \frac{\partial \lambda_i(x_i', \beta)}{\partial \beta_j} \right] \beta = \beta_0$$

and

$$\delta_i^0 = \beta_i - \beta_0^i$$

The Jacobian matrix of partial derivatives of $\lambda_i(x_i', \beta)$ with respect to the parameters, $\beta$ at the zeroth iteration is then

$$Z_0 = \begin{bmatrix} z_{11} & z_{12} & \cdots & z_{1p} \\ z_{21} & z_{22} & \cdots & z_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ z_{n1} & z_{n2} & \cdots & z_{np} \end{bmatrix} = Z_{ij}^0 \cdot n \times p.$$  

From (26), the residuals, $\epsilon_i$, now become

$$\epsilon_i = \lambda_i - \hat{\lambda}_i(x_i', \beta) - Z\delta.$$
3.5.1 Development of Dependent Variable, Y

Using (30), let \( \lambda_i - \lambda_i^0(x'_i\beta) \) be equal to \( Y_i \). We now rewrite the residuals

\[
e_i = Y_i - Z\delta,
\]

so that the column vector, \( Y \) of dependent \( Y \) becomes

\[
Y = \begin{bmatrix}
\lambda_1 - \lambda_1^0(x'_1\beta) \\
\lambda_2 - \lambda_2^0(x'_2\beta) \\
\vdots \\
\lambda_n - \lambda_n^0(x'_n\beta)
\end{bmatrix}.
\]

3.5.2 Development of Weighting Matrix, \( W \)

We know that \( Var(cX) = c^2 Var(X) \) so if we substitute in \( \tilde{\lambda} \) we have

\[
Var(\frac{1}{\lambda_i} \cdot \mu_i) = \left( \frac{1}{\lambda_i} \right)^2 Var(\mu_i) = \frac{\mu_i^2}{\lambda_i}.
\]

Therefore, each diagonal element of the weight matrix becomes

\[
w_{ii} = \frac{1}{Var(\lambda_i)} = \frac{T_i^2}{\mu_i} = \frac{T_i \cdot T_i}{\lambda_i} = \frac{T_i}{\lambda_i}.
\]

Because \( \lambda_i \) can be very small and \( T_i \) very large and it is better to divide by a larger number than a smaller one with a computer, we will choose the weight \( \frac{T_i^2}{\mu_i} \). The entire weighting matrix \( W \) is then

\[
W = \begin{bmatrix}
\frac{T_1^2}{\mu_1} & 0 & 0 & 0 \\
0 & \frac{T_2^2}{\mu_2} & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \frac{T_n^2}{\mu_n}
\end{bmatrix}.
\]

3.5.3 Defining Maximum Likelihood Solution Vector, \( \delta \)

Let \( V^{-1} = W \), and now substitute our new definition for residuals from (31) into (21) obtain the new sum of squares

\[
SS(\delta) = (Y - Z\delta)'W(Y - Z\delta) = WY'Y - (Z\delta)'WY - Y'ZW\delta + \delta'Z'WZ\delta.
\]

The transpose \( (Z\delta)'WY \) in (37) is equal to \( \delta'Z' \) because \( AB' = B'A' \). In addition, since \( (Z\delta)'WY \) is a 1 \times 1 matrix, or scalar, then its transpose \( Y'ZW\delta \) has the same value. Thus, (37) is transformed into

\[
SS(\delta) = WY'Y - 2\delta'Z'WY + \delta'Z'WZ\delta.
\]

The \( \delta' \) on the left sides of parameters in the right side of (38) drop out because if we set

\[
A = (Z'WY),
\]

and from Searle [25] know that

\[
\frac{\partial x'A}{\partial x} = A,
\]

6
and if we set
\[ A = (Z'WZ)\delta, \quad (41) \]
and from [25] know that
\[ \frac{\partial(x'Ax)}{\partial x} = 2Ax, \quad (42) \]
then the normal equation then becomes
\[ SS(\delta) = Z'WZ\delta + Z'WY. \quad (43) \]
In order to minimize (43) with respect to the \( \delta \) parameters, we must take the first partial derivative. This gives us
\[ \frac{\partial SS(\delta)}{\partial \delta} = -2Z'WY + 2(Z'WZ)\delta, \quad (44) \]
that provides us with the consistent equations
\[ (Z'WZ)\delta = Z'WY, \quad (45) \]
that has solution
\[ \delta = (Z'WZ)^{-1}Z'WY. \quad (46) \]
At the ith iteration, the values of the parameters are
\[ \beta_i = \beta_{i-1} + \delta. \quad (47) \]
Convergence is reached at the point when
\[ \{\beta_i - \beta_{(i-1)}\}/\beta_{(i-1)} < \epsilon, \quad (48) \]
where \( \epsilon \) is \( 10^{-5} \) by default in PIRLS.

### 3.6 Variance-covariance Matrix

The variance-covariance matrix, with parameter variance on the diagonal and covariances on the off-diagonals, at convergence is identical to the information matrix or inverse of the weighted dispersion matrix

\[ I^{-1}_\delta = (Z'WZ)^{-1} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1j} \\ \sigma_{12} & \sigma_2^2 & \cdots & \sigma_{2j} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1j} & \sigma_{2j} & \cdots & \sigma_j^2 \end{bmatrix}. \quad (49) \]

### 3.7 Standard Errors of Regression Coefficients

The standard errors of the regression coefficients are equivalent to \( \sqrt{\sigma_j^2} \). A Wald test for each parameter is simply \( \beta_j/\text{s.e.}(\beta_j) \) and indicates significance if the ratio is \( < -1.96 \) or \( > 1.96 \) at the \( \alpha = 0.05 \) level of significance.

### 3.8 Residuals and Regression Diagnostics

The error residual, \( e_i \) has mean 0 and measures the lack of concordance of the observed \( \mu_i \) against the fitted value \( \hat{\mu}_i \). The greater the value of \( e_i \), the worse the fit of the model. Pearson residuals, \( r_P \) are estimate as
\[ r_P = \frac{\mu_i - \hat{\mu}_i}{\mu}, \quad (50) \]
and provide a measure of the residual variation. The deviance residuals
\[ r_D = \mu_i \left[ \log \frac{\mu_i}{\hat{\mu}_i} \right] + (\hat{\mu}_i - \mu_i), \quad (51) \]
allow the investigator to determine the goodness of fit of the model, i.e., how well the observed data agree with the fitted values. Freeman and Tukey [26] introduced a variance stabilized residual for Poisson models given by

\[ r_{FT} = \sqrt{\mu + 1} - \sqrt{\mu + 1}. \] (52)

The leverage [23] for each observation, \( h_i \), provides information about the geometric distance between a given point \((z_{i1}, z_{i2}, \cdots, z_{ij})\) and the centroid of all points \((\bar{z}_{11}, \bar{z}_{22}, \cdots, \bar{z}_{ij})\) in the predictor space. Individual leverages, \( h_i \), are obtained from the diagonal of the “Hat” matrix given by

\[ H = \sqrt{\text{W} \text{Z}' \text{W}}^{-1} \text{Z}' \sqrt{\text{W}}. \] (53)

Because \( \sum h_i = \text{trace}(H) = p \), observations with high leverage can approach unity. A criterion introduced by Hoaglin and Welsch [27] for determining high leverage for an observation states that a point has high leverage if \( h_i > 2p/n \). Points with high leverage are located at a substantial distance from the fitted regression line, and act to “pull” the fitted line toward their location. Increases in the goodness of fit of a model can be obtained by refitting a model after the observation(s) with high leverage have been removed.

Cook [28] introduced the residual, \( \Delta(\beta)_{-ij} \), which measures the difference between \( \beta \) when the observation is included in the model and \( \beta \) when the observation is not in the model. The more positive the values of \( \Delta(\beta)_{-ij} \), the more influence an observation has when compared with other observations. Individual \( \Delta(\beta)_{-ij} \) residuals are obtained by the \( n \times p \) matrix \( \Delta(\beta)_{-ij} \), by the relationship

\[ \Delta(\beta)_{-ij} = \frac{-1}{1 - h_{ij}}. \] (54)

Finally, the standardized residual is used to measure the model error introduced by particular observation when its error residual has constant variance

\[ \hat{r}_i = \frac{e_i}{\sqrt{1 - h_i}}. \] (55)

Values for \( \hat{r}_i \) that exceed 1.96 tend to be outliers at the \( \alpha = 0.05 \) level of significance.

### 3.9 Goodness-of-Fit Statistics

The Pearson \( \chi^2 \) goodness of fit is

\[ \chi^2 = \sum_{i=1}^{n} r_p^2, \] (56)

and the deviance goodness of fit is

\[ D = \sum_{i=1}^{n} r_D, \] (57)

which is also \( \chi^2 \) distributed. If a model fits, its \( \chi^2 \) and \( D \) will be lower than the degrees of freedom \((n - p)\).

### 3.10 Structure

Program flow is outlined in Figure 1. The starting regression to obtain the initial values of the parameter vector, \( \beta_0 \), is performed by calling:

CALL INIT(NREG, NCOL, NROW, ISTART, START, B, G, RHO, IPRED, IDATA, INFILE, OUTFILE, IFLT)

For non-linear models, the user can specify starting values for the parameters and skip the multiplicative fit to obtain starting parameter estimates. The solution vector, \( \delta \), used to modify the original regression coefficients during each iteration are determined through calls of:

CALL PRED(NCOL, NROW, G, IDF, DEV, TPR, NREG, RHO, IPRED, ZTWZ1, ZTWY, ITER, WIN, YDIFF, OEXPE, OOBS, OFT, OP, XR, XIN, XT, IFLT)

Lastly, the leverages, deletion residuals, and standardized residuals are obtained by calling:

CALL FIT(WIN, YDIFF, XIN, XT, ZTWZ1, OEXPE, LEV, DRES, SRES, NCOL, NROW, IFLT)
\[ \begin{align*}
\text{Start} & \\
\text{Read } d_i, T_i, \text{ and } X & \\
\text{Estimate } \lambda_i = d_i/T_i, \ w_i = T_i & \\
\text{Additive} & : Y = \lambda \\
\text{Multiplicative} & : Y = \log(\lambda) \\
\text{Power} & : Y = \lambda^p \\
\text{Non-linear (optional)} & : Y = \log(\lambda) \\
\beta = (X'WX)^{-1}X'WY & \\
\text{Starting regression} & \\
\text{Iteratively Reweighted Least Squares} & \\
\delta_{i+1} = (Z'WZ)^{-1}Z'WY' & \quad \text{if } ||\delta_{i+1}|| > 10^{-5} \\
\text{Estimate } w_{ii}, z_{ij} & \\
\text{Estimate } Y_i = d_i - \lambda(x'\beta) T_i & \\
\text{Additive} & : \lambda(x'\beta) = x'\beta \\
\text{Multiplicative} & : \lambda(x'\beta) = \exp(x'\beta) \\
\text{Power} & : \lambda(x'\beta) = (x'\beta)^p \\
\text{Non-linear (example)} & : \lambda(x'\beta) = \beta_1 x_{i1} [1 - \{1 - e^{(-\beta_2 x_{i1})}\}^{\beta_3}] \\
\text{Stop when } ||\delta_{i+1}|| < 10^{-5} & \\
\end{align*} \]

Figure 1: Program flow in PIRLS.
The parameter, arrays, and variables used in PIRLS are defined below.

| Parameter  | Type       | Description                                                                 |
|------------|------------|-----------------------------------------------------------------------------|
| MP=4000    | Parameter  | Constant Fixed array size for maximum number of records                     |
| NP=28      | Parameter  | Constant Fixed array size for maximum number of variables                   |
| IP=1       | Parameter  | Constant Fixed array size for when 1 is needed                               |
| NIP=16     | Parameter  | Constant Fixed array when more than 1 is needed                               |
| INFILE     | Character  | Input Input filename and path                                               |
| OUTFILE    | Character  | Input Output filename and path                                               |
| NROW       | Integer    | Input Number of input records read                                           |
| NCOL       | Integer    | Input Number of parameters specified by user                                |
| ISTART     | Integer    | Input For non-linear models; 0 = user specified starting values, 1 = starting values estimated by multiplicative fit in subroutine INIT |
| START      | Real       | Input For non-linear models; starting values specified by user when ISTART=0 |
| IPRED      | Integer    | Input Number of parameters to be read in from input file                    |
| NREG       | Integer    | Input Type of regression model to be fit (1-additive, 2-multiplicative, 3-power, 4-non-linear) |
| IDATA      | Integer    | Input Write input data to output file, 1 = yes, 0 = no                      |
| ICOVAR     | Integer    | Input Write variance-covariance and correlation matrices to output file, 1 = yes, 0 = no |
| ICbELS     | Integer    | Input Write observed and fitted counts, \( d_i \), to output file, 1 = yes, 0 = no |
| IEX        | Integer    | Input Exponentiate regression coefficients for multiplicative models only, i.e., NREG=2 |
| IREG       | Integer    | Input Estimate standardized and deletion residuals and leverages, 1 = yes, 0 = no |
| RHO        | Real       | Input Power transform for power models, \( \rho \)                          |
| OBS(MP)    | Real       | Input Number of deaths or cases in subgroup (record), \( \mu_i \)            |
| Symbol  | Type   | Work          | Description                                                                 |
|---------|--------|---------------|------------------------------------------------------------------------------|
| T(MP)   | Real   | Input         | Number of person-years of follow-up in subgroup (record), T                   |
| X(MP,NP)| Real   | Work          | Design matrix of input independent variables, $X$                           |
| LAMBDA(MP) | Real | Work          | Observed rate, $\lambda_i$                                                  |
| LAMHAT(MP) | Real | Work          | Predicted rate, $\hat{\lambda}_i$                                          |
| EXPE(MP) | Real   | Work          | Expected deaths or cases in subgroup, $\mu_i = \lambda_i \times T$           |
| Y(MP)   | Real   | Work          | Error Residual, $Y_i = \mu_i - \hat{\mu}_i$                                |
| W(MP)   | Real   | Work          | Variance weights $W = T^2 / \text{Var}(\hat{\lambda}_i)$                   |
| Z(MP,NP) | Real   | Work          | Jacobian matrix of first partial derivatives, $Z$                            |
| ZTWZ(MP,NP) | Real | Work          | Information matrix, $(Z'WZ)$                                                 |
| ZTWZ1(NP,NP) | Real | Work          | Inverse of information matrix, $(Z'WZ)^{-1}$                                |
| ZTWY(NP,1) | Real | Work          | Score vector, $Z'WY$                                                         |
| FT(MP)  | Real   | Work          | Freeman-Tukey residuals, $r_{FT}$                                            |
| PR(MP)  | Real   | Work          | Pearson $\chi^2$ residuals, $r_P$                                           |
| DEV     | Real   | Work          | Deviance residuals, $r_D$                                                    |
| VARCOV(NP,NP) | Real | Work          | Variance-covariance matrix used for output, $(Z'WZ)$                         |
| G(NP,1) | Real   | Work          | Solution vectors, $\delta$                                                  |
| B(NP,1) | Real   | Work          | Regression coefficients updated at each iteration, $\beta_i$                |
| SE(NP)  | Real   | Work          | Standard error of $\beta_i$                                                 |
| CORR(NP,NP) | Real | Work          | Correlation matrix of coefficients                                            |
| LL(NP)  | Real   | Work          | Lower bound of 95% confidence limit for relative risk (multiplicative model only) |
| UL(NP)  | Real   | Work          | Upper bound of 95% confidence limit for relative risk (multiplicative model only) |
| ML(NP)  | Real   | Work          | Relative risk (multiplicative model only)                                    |
| LEV(MP,MP) | Real | Work          | Leverages from the Hat matrix, $h_i$                                         |
| DRES(MP,NP) | Real | Work          | Deletion residuals, $\Delta(\beta)_i$                                      |
| SRES(MP) | Real   | Work          | Standardized residuals, $\tilde{r}_i$                                      |
| IDF     | Integer| Work          | Degrees of freedom used for goodness of fit                                 |
| DEV     | Real   | Work          | Deviance goodness of fit, $D^2_{\hat{\phi},j}$                             |
| TPR     | Real   | Work          | Pearson $\chi^2$ goodness of fit $\chi^2_{\hat{\phi},j}$                  |
| IFLT    | Integer| Output        | Premature exit codes                                                         |

0=Convergence reached  
1=NCOL < 1  
2=NROW < 1  
3=Degrees of freedom, (NROW-NCOL), < 1  
4=$\sum |\beta_j| = 0$  
5=$\sum |G_j| = 0$  
6=Iterations > 100
3.11 Auxiliary Algorithms

MATMUL [29] performs matrix multiplication for the matrix manipulation. TRNPOS tranposes arrays as needed for matrix multiplication. MATINV calls SVDCMP based on [30] to invert the information matrix using singular value decomposition. ALG and GAMAIN are used for estimating tabled values of χ².

3.12 Restriction

All input parameters are checked for nullity and a fault message is returned if there is an illegal entry.

3.13 Time

A thorough investigation of absolute timing has not been performed; however, it should be noted that execution time is a function of the number of model parameters and effects to be considered.

3.14 Precision

The algorithm may be converted to double precision by making the following changes:

1. Change REAL to DOUBLE PRECISION in the algorithm.
2. Change the constants to double precision.
3. Change EXP to DEXP and ALOG to DLOG in all applicable routines.
4. Make appropriate changes in auxiliary routines MATMUL, TRNPOS, MATINV, SVDCMP, and in functions ALG and GAMAIN.

3.15 Application - Fitting a Multiplicative Model

3.15.1 Data Arrangement and Partial Derivatives

We shall enumerate the steps required for fitting a multiplicative model. Before we start the regression run, however, we must set up the data file and also determine the first partial derivatives of the rate ratio λ with respect to each parameter in the model.

1. **Arranging the data in an ASCII text file.** The first step is to arrange our data in an ASCII text file. The typical ASCII text file arrangement of data for a Poisson regression run using data in Table 1 is shown in the following:

   | Cases | T/1,000 | Age | Smoke |
   |-------|---------|-----|-------|
   | 2.    | 18.790  | 1   | 0     |
   | 12.   | 10.673  | 0   | 1     |
   | 28.   | 5.710   | 0   | 0     |
   | 28.   | 5.710   | 1   | 0     |
   | 31.   | 1.462   | 0   | 0     |
   | 32.   | 52.407  | 0   | 0     |
   | 104.  | 43.248  | 0   | 0     |
   | 206.  | 28.612  | 0   | 0     |
   | 186.  | 12.663  | 0   | 0     |
   | 102.  | 5.317   | 0   | 0     |

where the person-years in Table 1 are divided by 1,000 and the covariates denoting the age group are dummy coded (0,1) with a 1 to indicate that the rates are from a given cell, and whether or not the subgroup smoked.
2. **Determining Partial Derivatives of \( \lambda_i(x'_i|\beta) \) With Respect to Each \( \beta_j \).** Next we must write out the partial derivatives of the rate parameter with respect to each parameter (these will be used later in the algorithm). The partial derivatives are used in the Jacobian \( Z \) matrix to form the scores during each iteration. Thus, each partial derivative forms an element, \( z_{ij} \), in the Jacobian \( Z \) matrix. Given that the partial derivative of an exponentiated function \( u \) in the form \( e^u \) is \( e^u du \), the partial derivative of \( \lambda_i(x'_i|\beta) \) with respect to each parameter, \( \beta_j \) is as follows:

\[
z_{ij} = \frac{\partial \lambda_i(x'_i|\beta)}{\partial \beta_j} x_{ij},
\]

(58)

### 3.1.5.2 Steps in the Algorithm

The following section discusses the steps taken by the algorithm for performing iteratively reweighted least squares.

1. **Definition of the model.** The model we will fit is a log-linear model

\[
\lambda_i(x'_i|\beta) = e^{(\beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + \beta_6 x_{i6})},
\]

(59)

which is equivalent to

\[
\log(\lambda_i(x'_i|\beta)) = (\beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + \beta_6 x_{i6}).
\]

(60)

2. **Estimate column vector of dependent variables for the log-linear model.** The algorithm starts by reading in the data from the input file and then estimates the rate parameter, \( \lambda_i \) for each record as the ratio of number of *observed* deaths to the number of *observed* person-years. The log of \( \lambda_i \) is then calculated to obtain the dependent variable column vector

\[
Y = \begin{bmatrix}
\log(\frac{d_1}{d_0}) \\
\log(\frac{d_2}{d_3}) \\
\log(\frac{d_4}{d_5}) \\
\log(\frac{d_7}{d_6}) \\
\log(\frac{d_9}{d_8}) \\
\log(\frac{d_{10}}{d_9}) \\
\log(\frac{d_{12}}{d_{11}})
\end{bmatrix}.
\]

(61)

and when substituting in the data above we get

\[
Y = \begin{bmatrix}
\log(\frac{2}{18.506}) \\
\log(\frac{10.274}{28.407}) \\
\log(\frac{28.407}{28.407}) \\
\log(\frac{1.452}{28.407}) \\
\log(\frac{0.404}{43.248}) \\
\log(\frac{43.248}{43.248}) \\
\log(\frac{28.407}{28.407}) \\
\log(\frac{12.664}{12.664}) \\
\end{bmatrix}.
\]

(62)
3. Construction of the X matrix of independent variables. In addition to the estimation of the dependent variable \( Y \) column vector, the algorithm constructs the \( X \) matrix of independent variables from the independent variables read in from the input file as

\[
X = \begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 0 & 0 & 1 \\
0 & 1 & 0 & 0 & 0 & 1 \\
0 & 0 & 1 & 0 & 0 & 1 \\
0 & 0 & 0 & 1 & 0 & 1 \\
0 & 0 & 0 & 0 & 1 & 1 \\
\end{bmatrix}.
\]

(63)

4. “Starting Regression” - log-linear regression to estimate initial values of parameters. Once the \( Y \) and \( X \) matrices are constructed, the algorithm fits the log-linear model

\[
\log(\lambda_i(x'_i\beta)) = (\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6),
\]

(64)

which in matrix terms is given as

\[
\beta = (X'WX)^{-1}X'WY,
\]

(65)

where diagonal elements of \( W \) are set equal to \( T \). Figure 1 shows the program flow in PIRLS for the various models. Table 3 shows the link functions for \( \lambda \) for each starting regression and the linear predictor used for estimating \( \lambda_i(x'_i\beta) \) during subsequent iterations.

| Model          | Link function for \( \lambda \) | Form of predictor for \( \lambda_i(x'_i\beta) \) |
|---------------|---------------------------------|-----------------------------------------------|
| Additive      | \( \lambda_i \)                 | \( \lambda_i(x'_i\beta) = (x'_i\beta) \)     |
| Multiplicative| \( \log(\lambda) \)              | \( \lambda_i(x'_i\beta) = \exp(x'_i\beta) \) |
| Power         | \( \lambda^p \)                 | \( \lambda_i(x'_i\beta) = (x'_i\beta)^p \)   |
| Non-linear (optional)* | \( \log(\lambda) \) | \( \lambda_i(x'_i\beta) = \exp(x'_i\beta) \) |

* For non-linear models, the user can specify starting values for parameters, rather than using the default multiplicative fit to obtain starting values.

5. Estimation of predicted rate parameter. After the first regression or iteration, the regression coefficients, \( \beta_j \), are used to estimate predicted rates from the linear predictor

\[
\lambda_i(x'_i\beta) = e^{\sum_{j=1}^p x_{ij}\beta_j} = e(x'_i\beta).
\]

(66)

6. Estimation of expected number of deaths. The predicted rates are then multiplied by the observed person-years in each record to give the expected number of deaths as

\[
\hat{\mu}_i = d_i(x'_i\beta) = \lambda_i(x'_i\beta)T_i.
\]

(67)

7. Estimation of dependent variable. The expected number of deaths, \( \hat{\mu}_i \), are then subtracted from the observed deaths, \( \hat{\mu}_i \), to give the dependent variable

\[
Y_i = \mu_i - \hat{\mu}_i = d_i - \lambda_i(x'_i\beta)T_i.
\]

(68)

8. Construction of column vector \( Y \). The column vector \( Y \) is then equal to \([Y_i] \).

14
Table 4: Survival data for splenic stem cell colony formation after irradiation and transplantation (see ref. [31]).

| Colonies, $\mu$ | Trials, $T_i$ | Cell concentration, $x_{i1}$ | Dose(Sv), $x_{i2}$ |
|-----------------|---------------|-------------------------------|------------------|
| 60.0            | 6.0           | 1.25                          | 0.00             |
| 66.0            | 7.0           | 1.75                          | 0.96             |
| 46.0            | 4.0           | 3.00                          | 1.92             |
| 82.0            | 9.0           | 7.20                          | 2.88             |
| 105.0           | 11.0          | 24.00                         | 4.32             |
| 123.0           | 15.0          | 75.00                         | 5.76             |
| 12.0            | 4.0           | 120.00                        | 6.72             |

9. **Estimation of the partial derivatives for each row vector of the data matrix $X$.** For each record in the data matrix, $X$, estimate the value of the partial derivatives using the current parameter values, $\beta_j$. For our log-linear model in (65), we get

$$ z_{ij} = \frac{\partial \ln(\mu_j)}{\partial \beta_j} = e^{(\beta - \beta_j)} x_{ij} . $$

10. **Construction of the Jacobian matrix of partial derivatives.** Now the algorithm constructs the $n \times p$ Jacobian matrix $Z$ composed of the first partial derivatives estimated above, which appears as $Z = [z_{ij}]$.

11. **Construction of the inverse variance weighting matrix.** Since the Poisson variance of $\hat{\mu}_i$ changes at each iteration, we must weight the dispersion matrix with weights defined by $T_i^2/\hat{\mu}_i$ given as $W = \text{diag}[w_{ii}]$, where $w_{ii} = \frac{T_i^2}{\hat{\mu}_i}$.

12. **Estimation the solution vector at each iteration.** The maximum likelihood solutions at the next and subsequent iterations are obtained through the matrix manipulation

$$ \delta = (Z'WZ)^{-1}Z'WY . $$

13. **Updating regression coefficients.** At each iteration, the new solution vectors are added to the previous parameter estimates given as

$$ \beta_i = \beta_{i-1} + \delta . $$

14. **Repeat steps 5-13 above until convergence is reached.** The above steps (5 through 13) are repeated until the euclidean norm of the score vector is below some non-negative value

$$ \{ \beta_i - \beta_{i-1} \}/\beta_{i-1} < \epsilon , $$

where $\epsilon = 10^{-5}$ by default in PIRLS. At this point, convergence is reached at a global maximum.

### 3.16 Application - Fitting a Non-linear Model

For this example, we will fit a non-linear model to estimate the number of stem cell colonies in spleens of laboratory mice into which irradiated cells were transplanted [31]. Table 4 shows the colonies formed, $\mu$, number of trials, $T_i$, cell concentration, $x_{i1}$, and radiation dose (Sv), $x_{i2}$, arranged in tabular notation.

The non-linear model we will fit to obtain maximum likelihood estimates of colonies formed is of the form

$$ \lambda_i = \beta_1 x_{i1} [1 - (1 - e^{(-\beta_2 x_{i2})})^{\beta_3}] , $$

where $x_{i1}$ is the concentration of transplanted cells, $x_{i2}$ is the radiation dose (Sv), and $\beta_1$, $\beta_2$, and $\beta_3$ are the parameters of interest.
In order to build the Jacobian matrix, \( Z \), we need to take the first partial derivatives of \( \lambda_i \) with respect to \( \beta_1 \), \( \beta_2 \), and \( \beta_3 \), i.e., \( \partial\lambda_i/\partial \beta_1 \), \( \partial\lambda_i/\partial \beta_2 \), and \( \partial\lambda_i/\partial \beta_3 \). Since there are three \( \beta \) coefficients, the first column entries for row \( i \) of \( Z \), i.e., \( z_{i1} \), will be estimated by use of equation

\[
\frac{\partial \lambda_i}{\partial \beta_1} = x_{i1} [1 - \{1 - e^{(-\beta_2 x_{i2})}\}^{\beta_3}].
\] (74)

The first partial derivative in the 2nd column entry of row \( i \) of \( Z \), i.e., \( z_{i2} \) is estimated as

\[
\frac{\partial \lambda_i}{\partial \beta_2} = -\beta_1 x_{i1} \{1 - e^{(-\beta_2 x_{i2})}\}^{\beta_3} \beta_3 x_{i2} \frac{e^{-\beta_2 x_{i2}}}{1 - e^{-\beta_2 x_{i2}}}. \] (75)

The third column of \( Z \) has elements \( z_{i3} \), which are numerically estimated as

\[
\frac{\partial \lambda_i}{\partial \beta_3} = -\beta_1 x_{i1} \{1 - e^{(-\beta_2 x_{i2})}\}^{\beta_3} \ln(1 - e^{-\beta_2 x_{i2}}). \] (76)

When fitting this model, we chose to specify starting values of \( \beta_1 = 7.6364 \), \( \beta_2 = 0.9341 \), and \( \beta_3 = 7.6364 \) from [3], rather than use the default multiplicative fit to obtain starting values. Input and output for this model using the data in Table 4 are shown in the Appendix.

### 3.17 Input

During run-time, the user must respond to the following queries concerning each run:

- The type of model to be fit (1-additive, 2-multiplicative, 3-power, and 4-non-linear)
- For additive models, the power, \( \rho \), which ranges from 1 for purely additive models to 0 for purely multiplicative models
- The number of independent (input) variables to be read in from the input file
- The number of parameters to estimate
- For non-linear models, optional starting values for parameters
- For multiplicative models, the regression coefficients are to be exponentiated to obtain RR and its 95% CI
- Whether or not the variance-covariance and correlation matrices are written to the output file
- Whether or not the input data are written to the output file
- Whether or not the observed and fitted count data (deaths or cases) are written to the output file
- Whether or not the leverages, and standardized and deletion residuals are written to the output file

### 4 USER INPUT DATA FILE

PIRLS uses one input data file (ASCII) whose name is specified by the user at run-time. The input file contains three types of information

1. Run title (text) in record 1
2. FORTRAN format statement (in parentheses) in record 2 needed for reading the input data in later records
3. Input data, i.e., cases, person-years, and independent variables in all subsequent records

The examples in the Appendix give listings of input files.

### 5 SAMPLE INPUT/OUTPUT CASES

The output for all runs is in tabular form in the ASCII text file whose name is specified by the user at run-time. Four examples using an input file are given in the Appendix.
Table 5: Names and descriptions of each PIRLS file.

| Filename  | Description       | Format       |
|-----------|------------------|--------------|
| INFILE*   | Input filename   | ACSII text   |
| OUTFILE*  | Output filename  | ACSII text   |
| PIRLS.FOR | FORTRAN source code | ASCII text |
| PIRLS.EXE | Executable file  | FORTRAN      |

* Input and output filenames are specified by user at run-time.

6 FILENAMES

Table 5 lists the names of the files that were used to program, link and execute PIRLS. As one notices, the source code is an ASCII text file and the object and executable files have been compiled with Microsoft FORTRAN Powerstation 4.0. There is one required input data file, whose name is specified at run-time.

7 AVAILABILITY

The program and all of its subroutines are available from the Journal of Statistical Software free of charge at http://www.stat.ucla.edu/journals/jss/

8 ACKNOWLEDGEMENTS

The author is grateful to the reviewers for helpful comments. Work on this algorithm was performed under NASA contract NAS-9 18236.

References

[1] Frome, E.L., Kutner, M.H., Beauchamp, J.J. (1973) Regression analysis of Poisson-distributed data. JASA. 68:935-940.
[2] Frome, E.L. (1983) The analysis of rates using Poisson regression methods. Biometrics. 39:665-674.
[3] Frome, E.L. PREG: A computer program for Poisson regression analysis. Oak Ridge Associated Universities Report ORAU-178 (1981). Oak Ridge National Laboratory. Oak Ridge(TN), ORAU.
[4] Muirhead, C.R., Darby, S.C. (1987) Modelling the relative and absolute risks of radiation-induced cancers. J. R. Stat. Soc. A. 150:83-118.
[5] Breslow, N.E., Day, N.E. (1987) Statistical Methods in Cancer Research. Volume II: The design and analysis of cohort studies. Chapter 4. IARC Scientific Publication No. 82. Lyon, IARC.
[6] Pierce, D.A., Shimizu, Y., Preston, D.L., Vaeth, M., Mabuchi, K. (1996) Studies of mortality of atomic bomb survivors. Report 12, Part I. Cancer: 1950-1990. Radiat. Res. 146:1-27.
[7] Thompson, D.E., Mabuchi, K., Ron, E., Soda, M., Tokunaga, M., Ochikubo, S., Sugimoto, S., Ikeda, S., Terasaki, M., Izumi, S. (1994) Cancer Incidence in Atomic Bomb Survivors. Part II: Solid Tumors, 1958-1987. Radiat. Res. 137:S17-S67.
[8] Preston, D.L., Kusimi, S., Tomonaga, M., Izumi, S., Ron, S., Kuramoto, A., Kamada, N., Dohy, H., Matsui, T., Nonaka, H., Thompson, D.E., Soda, M., Mabuchi, K. (1994) Cancer Incidence in Atomic Bomb Survivors. Part III: Leukemia, Lymphoma and Multiple Myeloma, 1950-87. Radiat. Res. 137, S68-S97; 1994.

[9] Cardis, E., Gilbert, E.S., Carpenter, L., Howe, G., Kato, I., Armstring, B.K., Beral, V., Cowper, G., Douglas, A., Fix, J., Fry, S.A., Kaldor, J., Lavé, C., Salmon, L., Smith, P.G., Voelz, G.L., Wiggs, L.D. (1995) Effects of low doses and low dose rates of external ionizing radiation: Cancer mortality among nuclear industry workers in three countries. Radiat. Res. 142: 117-132.

[10] Ron, E., Lubin, J.H., Shore, R.E., Mabuchi, K., Modan, B., Pottern, L.M., Schneider, A.B., Tucker, M.A., Boice, J.D. Jr. (1995) Thyroid cancer after exposure to external radiation: a pooled analysis of seven studies. Radiat. Res. 141:259-277.

[11] Kerber, R.A., Till, J.E., Simon, S.E., Lyon, J.L., Thomas, D.C., Preston-Martin, S., Rallison, M.L., Lloyd, R.D., Stevens, W.S. (1993) A cohort study of thyroid disease in relation to fallout from nuclear weapons testing. J. Am. Med. Assoc. 270:2076-2082.

[12] Peterson, L.E., Schull, W.J., Davis, B.R., Buffle, P.A. (1994) Information Bias and Lifetime Mortality Risks of Radiation-Induced Cancer. U. S. NRC NUREG Report GR-0011. Nuclear Regulatory Commission, Washington, D.C.

[13] Spostto, R., Preston, D.L, Shimizu, Y., Mabuchi, K. (1992) The effect of diagnostic misclassification on non-cancer and cancer mortality dose-response in A-bomb survivors. Biometrics. 48:605-617.

[14] Whittemore, A.S., Gong, G. (1991) Poisson regression with misclassified counts: application to cervical cancer mortality rates. Appl. Stat. 40: 81-93.

[15] Frome, E.J., DuFrain, R.J. (1986) Maximum likelihood estimation for cytogenetic dose-response curves. Biometrics. 42:73-84.

[16] Ripley, B.D., Venables, W.N. (1994) Modern Applied Statistics with S-PLUS. New York, Springer-Verlag.

[17] Tierney, L. (1990) Lisp-Stat: An object-oriented environment for statistical computing and dynamic graphics. New York, John Wiley and Sons.

[18] STATA (1997) - Poisson Regression, User’s Guide. College Station(TX), Stata Press.

[19] Presto, D.L. and Pierce, D.A. (1997) AMFIT: A program for parameter estimation in additive and multiplicative rate models with grouped survival data - methods, models, and examples. AMFIT User’s Guide. Hirosoft International Corporation. Seattle(WA), Hirosoft International Corporation.

[20] EGRET (1997) - Poisson Regression, User’s Guide. Statistics and Epidemiology Research Corporation. Seattle(WA), SERC.

[21] Doll, R., Hill, A.B. (1966) Mortality of British doctors in relation to smoking: Observations on coronary thrombosis. Natl. Cancer Inst. Monogr. 19:205-268.

[22] Draper, N.R., Smith, H. (1981) Applied Regression Analysis. New York, John Wiley and Sons.

[23] McCullagh, P., Nelder, J.A. (1983) Generalized Linear Models. London, Chapman and Hall.

[24] Aranda-Ordaz, F.J. (1983) An extension of the proportional-hazards model for grouped data. Biometrics. 39:109-117.

[25] Searle, S.R. (1982) Matrix Algebra Useful for Statistics. New York, John Wiley and Sons.

[26] Freeman, M.F., Tukey, J.W. (1950) Transformations related to the angular and the square root. Annals of Stat. 21:607-611.

[27] Hoaglin, D.C., Welsch, R.E. (1978) The Hat matrix in regression and ANOVA. Am. Statistician. 32:17-22.

[28] Cook, R.D. (1977) Detection of influential observations in linear regression. Technometrics. 19:15-18.

[29] Hieberger, R. (1989) Computation for the analysis of designed experiments. New York, John Wiley and Sons.

[30] Press, W.H., Flannery, B.P., Teukolsky, S.A., Vetterling, W.T. (1989) Numerical Recipes: The Art of Scientific Computing. New York, Cambridge University Press.

[31] Till, J.E., McCulloch, E.A. (1961) A direct measurement of radiation sensitivity of normal mouse bone marrow cells. Radiat. Res. 14:213-222.
9 APPENDIX: SAMPLE INPUT/OUTPUT CASES

Example 1: Additive (linear) model using data in Table 1 adapted from [21]. The input data file for this run is distribution file EXAMP1-3.DAT and is listed below:

```
PIRLS
(F4.0,F8.3,6F3.0)
   2. 18.790 1. 0. 0. 0. 0.
   12. 10.673 0. 1. 0. 0. 0.
   28. 5.710 0. 0. 1. 0. 0.
   30. 2.585 0. 0. 0. 1. 0.
   31. 1.462 0. 0. 0. 0. 1.
   32. 52.407 1. 0. 0. 0. 1.
  104. 43.248 0. 1. 0. 0. 1.
  206. 28.612 0. 0. 1. 0. 1.
  186. 12.663 0. 0. 0. 1. 1.
  102. 5.317 0. 0. 0. 0. 1.
```

The screen queries to be entered by the user are as follows:

```
ENTER THE NAME OF INPUT FILE:
EXAMP1-3.DAT

ENTER THE NAME OF INPUT FILE:
EXAMPLE1.OUT

ENTER 1-ADD 2-MULT 3-POWER 4-NON-LINEAR
1

ENTER RHO: 1 - ADDITIVE  - OTHER
1

ENTER NUMBER OF PREDICTOR (INPUT) VARIABLES
6

NUMBER OF PARAMETERS TO BE FITTED?
6

PRINT INFORMATION AND CORR MATRICES? 0-NO 1-YES
1

PRINT INPUT DATA? 0-NO 1-YES
1

PRINT FITTED CELLS? 0-NO 1-YES
1

PERFORM REGRESSION DIAGNOSTICS? 0-NO 1-YES
1
```

The output file contains:

```
TITLE: PIRLS
```
MODEL SELECTED: ADDITIVE

INPUT DATA FORMAT: (F4.0,F8.3,6F3.0)

2. 18.790 1. 0. 0. 0. 0. 0.
12. 10.673 0. 1. 0. 0. 0. 0.
28. 5.710 0. 0. 1. 0. 0. 0.
28. 2.585 0. 0. 0. 1. 0. 0.
31. 1.462 0. 0. 0. 0. 1. 0.
32. 52.407 1. 0. 0. 0. 0. 1.
104. 43.248 0. 1. 0. 0. 0. 1.
206. 28.612 0. 0. 1. 0. 0. 1.
186. 12.663 0. 0. 0. 1. 0. 1.
102. 5.317 0. 0. 0. 0. 1. 1.

ITERATION: 1 DEVIANCE: 86.7977
ITERATION: 2 DEVIANCE: 89.7618
ITERATION: 3 DEVIANCE: 93.6565
ITERATION: 4 DEVIANCE: 98.2919
ITERATION: 5 DEVIANCE: 22.3763
ITERATION: 6 DEVIANCE: 7.7339
ITERATION: 7 DEVIANCE: 7.4384
ITERATION: 8 DEVIANCE: 7.4331
ITERATION: 9 DEVIANCE: 7.4330
ITERATION: 10 DEVIANCE: 7.4330
ITERATION: 11 DEVIANCE: 7.4330
ITERATION: 12 DEVIANCE: 7.4330
ITERATION: 13 DEVIANCE: 7.4330

ITERATIONS: 13

REC OBS EXP FREEMAN-TUKEY PEARSON
1 2.0 1.6  .4403  .1113
2 12.0 17.5 -1.3592  1.7346
3 28.0 36.0 -1.3638  1.7751
4 28.0 35.0 -1.1909  1.3856
5 31.0 28.0  .5896  .3156
6 32.0 35.4  -.5343  .3202
7 104.0 96.5  .7722  .5822
8 206.0 197.3  .6327  .3874
9 186.0 178.7  .5558  .2952
10 102.0 105.1  -.2763  .0895

B COEFFICIENT STD. ERROR WALD SCORE
1  .0841  .0661  1.2724  .00013366
2  1.6407  .2179  7.5310  -.00000497
3  6.3035  .4565 13.8080  -.00000113
4 13.5241  .9642 14.0270  -.00000030
5 19.1696 1.7045 11.2466  -.00000012
6  .5907  .1255  4.7055  -.00000556

POWER: 1.00

GOODNESS-OF-FIT TESTS

| STATISTIC   | ESTIMATE | D.F. | PROB. |
|-------------|----------|------|-------|
| CHI-SQUARE  | 6.997    | 4    | .1361 |
| DEVIANCE    | 7.433    | 4    | .1147 |

VARIANCE-COVARIANCE MATRIX OF COEFFICIENTS.

| B1  | B2   | B3   | B4   | B5   | B6   |
|-----|------|------|------|------|------|
| B1  | .004 | .003 | .003 | .003 | .003 |
| B2  | .003 | .047 | .010 | .010 | .009 |
| B3  | .003 | .010 | .208 | .011 | .930 |
| B4  | .003 | .010 | .011 | .010 | 2.905|
| B5  | .003 | .009 | .010 | .010 | .012 |
| B6  | -.004| -.012| -.013| -.012| .016 |

CORRELATION MATRIX OF COEFFICIENTS.

| B1  | B2   | B3   | B4   | B5   | B6   |
|-----|------|------|------|------|------|
| B1  | 1.000| .111 | .053 | .028 | -.490|
| B2  | .211 | 1.000| .046 | .025 | -.431 |
| B3  | .111 | .097 | 1.000| .013 | -.226 |
| B4  | .053 | .046 | .024 | 1.000| -.107 |
| B5  | .028 | .025 | .013 | .006 | 1.000|
| B6  | -.490| -.431| -.226| -.107| 1.000|

TRACE OF THE HAT MATRIX AND ADJUSTED RESIDUALS.

| RECORD | LEVERAGE | ADJ. RESID. |
|--------|----------|-------------|
| 1      | .9763    | 2.1668      |
| 2      | .3088    | -1.5841     |
| 3      | .1888    | -1.4792     |
| 4      | .1777    | -1.2981     |
| 5      | .2216    | .6367       |
| 6      | .9318    | -2.1671     |
| 7      | .7680    | 1.5841      |
| 8      | .8229    | 1.4792      |
On the first page of the output, we see that the model is indeed additive, we also see the input data and the value of the deviance goodness of fit at each iteration. The number of observed and expected deaths are given and the Freeman-Tukey and Pearson residuals as well. At convergence, the coefficient for smoking (No. 6) is 0.5907. Thus, smoking adds 0.59 deaths per 1,000 person-years (or 59/100,000) of follow-up at each age. However, since the $\chi^2$ and $D$ goodness of fit statistics are greater than 4 degrees of freedom, we conclude that the model does not fit the data well. Table 6 shows the comparison between the modeled baseline rates and modeled baseline rates to which 0.59/1,000 was added.

### Table 6: Interpretation of absolute risks from the additive model.

| Age group, i | Baseline rate, $\lambda_{i0}(x_{i0}\beta)$ per 1,000 $T$ | Smoker rate, $\lambda_{i1}(x_{i1}\beta + \beta_{\text{smoke}})$ per 1,000 $T$ |
|--------------|-------------------------------------------------------------|-------------------------------------------------------------|
| 35-44        | 0.0841                                                      | 0.6748                                                      |
| 45-54        | 1.6407                                                      | 2.2314                                                      |
| 55-64        | 6.3035                                                      | 6.8942                                                      |
| 65-74        | 13.5241                                                     | 14.115                                                      |
| 75-84        | 19.1696                                                     | 19.7603                                                     |

Example 2: Multiplicative (log-linear) model using data in Table 1 adapted from [21]. The input data file for this run is the same as that in Example 1.

The screen queries to be entered by the user are as follows:

**ENTER THE NAME OF INPUT FILE:**

EXAMP1-3.DAT

**ENTER THE NAME OF INPUT FILE:**

EXAMPLE2.OUT

**ENTER** 1-ADD 2-MULT 3-POWER 4-NON-LINEAR 2

2
ENTER NUMBER OF PREDICTOR (INPUT) VARIABLES
6

NUMBER OF PARAMETERS TO BE FITTED?
6

EXPONENTIATE COEFFICIENTS? 0-NO 1-YES
1

PRINT INFORMATION AND CORR MATRICES? 0-NO 1-YES
0

PRINT INPUT DATA? 0-NO 1-YES
0

PRINT FITTED CELLS? 0-NO 1-YES
0

PERFORM REGRESSION DIAGNOSTICS? 0-NO 1-YES
0

The output file contains:

TITLE: PIRLS

MODEL SELECTED: MULTIPLICATIVE

INPUT DATA FORMAT: (F4.0,F8.3,6F3.0)

ITERATION:  1  DEVIANCE:  53.9002
ITERATION:  2  DEVIANCE:  17.9280
ITERATION:  3  DEVIANCE:  12.1932
ITERATION:  4  DEVIANCE:  12.1324
ITERATION:  5  DEVIANCE:  12.1324

| B  | COEFFICIENT | STD. ERROR | WALD  | SCORE          |
|----|-------------|------------|-------|----------------|
| 1  | -1.0116     | .1918      | -5.2751| .00000207      |
| 2  | .4724       | .1304      | 3.6236 | .00000969      |
| 3  | 1.6159      | .1147      | 14.0944| -.00002298     |
| 4  | 2.3389      | .1162      | 20.1343| .00001905      |
| 5  | 2.6885      | .1250      | 21.5110| .00001919      |
At convergence, the coefficient for smoking (No. 6) is 0.3545. Since we exponentiated the regression coefficients, we can interpret the RR for smoking and the 95\% confidence interval as being 1.43(1.15,1.76). This implies that the risk of coronary death in smokers is 1.43 times the age-specific rates of the non-smokers. The goodness of fit statistics for this run are greater than those generated in Example 1, indicating that the multiplicative model fits worse when compared with the additive model. Table 7 shows the comparison between modeled baseline rates and the modeled baseline rates multiplied by the RR of 1.43.

Table 7: Interpretation of relative risks from the multiplicative model.

| Age group, \(i\) | Baseline rate, \(\lambda_{0i}(x_{0i}\beta)\) (\(e^{\beta_j}\)) per 1,000 \(T\) | Smoker rate, \(\lambda_{1i}(x'_{0i}\beta)\) (\(\lambda_{0i}(x'_{0i}\beta) \times e^{\beta_{\text{smoke}}}\)) per 1,000 \(T\) |
|-----------------|--------------------------------|--------------------------------|
| 35-44           | 0.36                          | 0.5148                          |
| 45-54           | 1.60                          | 2.28                            |
| 55-64           | 5.03                          | 7.192                           |
| 65-74           | 10.37                         | 14.83                           |
| 75-84           | 14.71                         | 21.04                           |

Example 3: Power model using data in Table 1 adapted from [21]. The input data file for this run is the same as that in Example 1.

The screen queries to be entered by the user are as follows:

ENTER THE NAME OF INPUT FILE:
EXAMP1-3.DAT

ENTER THE NAME OF INPUT FILE:
EXAMPLE3.OUT

ENTER 1-ADD 2-MULT 3-POWER 4-NON-LINEAR
3

ENTER NUMBER OF PREDICTOR (INPUT) VARIABLES
6

NUMBER OF PARAMETERS TO BE FITTED?
6

EXPONENTIATE COEFFICIENTS? 0-NO 1-YES
This run fitted ten models with $\rho$ varying from one to zero, i.e., going from a purely additive link, $\rho = 1$, to a purely multiplicative link, $\rho = 0$. The model equation is $y_i = (x_0\beta)^\rho$ where $\rho$ is the power function. By inspecting the fits of the models as a function of $\rho$, one can see that for these example data, the lowest deviance is obtained with $\rho$ between 0.5 to 0.6. A single run can now be made to fit the model with $\rho = 0.55$ by following the directions given in Example 1, but by specifying 0.55 for $\rho$ instead of one. Table 8 lists the modeled baseline rates, $\beta_j^{1/0.55}$, and modeled baseline rates to which $\beta_{\text{smoke}}^{1/0.55}$ was added.

Table 8: Interpretation of relative risks from the power model.

| Age group $i$ | Baseline rate, $\lambda_{i0}(x_{i0}\beta)$ | Smoker rate, $\lambda_{i1}(x_{i0}\beta)$ |
|--------------|-------------------------------------------|------------------------------------------|
|              | ($\beta_j^{1/0.55}$) per 1,000 $T$        | ($\lambda_{i0}(x_{i0}\beta) + \beta_{\text{smoke}}^{1/0.55}$) per 1,000 $T$ |
| 35-44        | 0.096                                     | 0.3742                                   |
| 45-54        | 1.218                                     | 1.494                                    |
| 55-64        | 5.123                                     | 5.400                                    |
| 65-74        | 11.649                                    | 11.925                                   |
| 75-84        | 17.081                                    | 17.357                                   |

Example 4: Non-linear model using cell concentrations, and radiation doses to estimate number of splenic stem cell colonies formed after transplantation into laboratory mice. The input data (see Table 4) for this run is provided in distribution file EXAMPLE4.DAT and is listed below:

Till and McCulloch (1961)
(F4.0,F4.0,F7.2,F6.2)

60. 6. 1.25 .00
66. 7. 1.75 .96
46. 4. 3.00 1.92
82. 9. 7.20 2.88
105.11. 24.00 4.32
123.15. 75.00 5.76
12.4. 120.00 6.72

The screen queries to be entered by the user are as follows:

ENTER THE NAME OF INPUT FILE:
EXAMPLE4.DAT

ENTER THE NAME OF INPUT FILE:
Results in the output file EXAMPLE4.OUT (filename specified above) indicate that 2 iterations were required before convergence was reached. Although the deviance and chi-square GOF values were greater than the 4 d.f., the Wald statistics for the parameters were statistically significant. The scores for all coefficients were near zero. The regression parameters, standard errors (with the exception of s.e.(\(\hat{\beta}_1\))), correlation matrix, chi-square and deviance GOFs were identical to results in Frome [3].

**TITLE:** Till and McCulloch (1961)

**MODEL SELECTED:** NON-LINEAR

**INPUT DATA FORMAT:** (F4.0,F4.0,F7.2,F6.2)
STARTING VALUES FOR PARAMETERS

PARAMETER STARTING VALUE

1 7.6364000
2 .9341000
3 2.8924000

ITERATION: 1  DEVIANCE:  8.0174
ITERATION: 2  DEVIANCE:  8.0174

ITERATIONS: 2

| REC | OBS | EXP | FREEMAN-TUKEY | PEARSON |
|-----|-----|-----|---------------|---------|
| 1   | 60  | 57  | .3874         | .1298   |
| 2   | 66  | 73  | -.8079        | .6713   |
| 3   | 46  | 37  | 1.3496        | 1.9260  |
| 4   | 82  | 91  | -.9414        | .8942   |
| 5   | 105 | 101 | .3807         | .1296   |
| 6   | 123 | 113 | .8536         | .7194   |
| 7   | 12  | 19  | -1.9041       | 3.1247  |

B COEFFICIENT STD. ERROR WALD SCORE

|    | B1   |     |     |     |     |
|----|------|-----|-----|-----|-----|
| 1  | 7.6364 | .1186 | 64.3725 | .00030114 |
| 2  | .9341  | .0399 | 23.4229 | -.00096028 |
| 3  | 2.8924 | .7476 | 3.8688  | .00006890 |

POWER: .00

GOODNESS-OF-FIT TESTS

STATISTIC ESTIMATE D.F. PROB.

| STATISTIC | ESTIMATE | D.F. | PROB. |
|-----------|----------|------|-------|
| CHI-SQUARE | 7.595    | 4    | .1076 |
| DEVIANCE   | 8.017    | 4    | .0909 |

VARIANCE-COVARIANCE MATRIX OF COEFFICIENTS.

|    | B1  | B2  | B3  |
|----|-----|-----|-----|
| B1 | .014|     |     |
| B2 | -.002| .002|     |
CORRELATION MATRIX OF COEFFICIENTS.

|     | B1   | B2   | B3   |
|-----|------|------|------|
| B1  | 1.000| -0.343| -0.741|
| B2  | -0.343| 1.000| 0.853|
| B3  | -0.741| 0.853| 1.000|

TRACE OF THE HAT MATRIX AND ADJUSTED RESIDUALS.

| RECORD | LEVERAGE | ADJ. RESID. |
|--------|----------|-------------|
| 1      | 0.8060   | 0.8181      |
| 2      | 0.3663   | -1.0293     |
| 3      | 0.2251   | 1.5766      |
| 4      | 0.4551   | -1.2811     |
| 5      | 0.2954   | 0.4289      |
| 6      | 0.6358   | 1.4055      |
| 7      | 0.2162   | -1.9966     |

DELETION RESIDUALS OR APPROXIMATE CHANGE IN COEFFICIENTS AFTER DELETION OF EACH RECORD.

| RECORD | B1   | B2   | B3   |
|--------|------|------|------|
| 1      | -0.1978 | 0.0228 | 0.9235 |
| 2      | 0.0581  | 0.0141 | 0.0336 |
| 3      | 0.0065  | -0.0276 | -0.4434 |
| 4      | -0.0339 | 0.0336 | 0.6643 |
| 5      | 0.0060  | -0.0010 | -0.0749 |
| 6      | -0.0201 | 0.0499 | 0.4879 |
| 7      | 0.0231  | -0.0354 | -0.4386 |