Robust Regression Analysis for Non-Normal Situations under Symmetric Distributions Arising In Medical Research

S S. Ganguly

Sultan Qaboos University, Muscat, Oman, ganguly@squ.edu.om

Follow this and additional works at: http://digitalcommons.wayne.edu/jmasm

Part of the Applied Statistics Commons, Social and Behavioral Sciences Commons, and the Statistical Theory Commons

Recommended Citation
DOI: 10.22237/jmasm/1398918480
Available at: http://digitalcommons.wayne.edu/jmasm/vol13/iss1/29
Robust Regression Analysis for Non-Normal Situations under Symmetric Distributions Arising In Medical Research

S. S. Ganguly
Sultan Qaboos University
Muscat, Oman

Introduction

Often in medicine, a relationship is established between a response variable \( y \), which depends on the \( r \) covariates \( x_1, x_2, \ldots, x_r \), which are independent of each other, so that, in total, there may be \( (r + 1) \) variables. In classical regression model, the response variable \( y \) is treated as a random variable whose mean depends upon fixed variables of the \( x_i \)’s. The mean is assumed to be linear function of the regression coefficients \( a, \beta_1, \beta_2, \ldots, \beta_r \).

The linear regression model also arises in a different setting. Suppose all the variables \( y, x_1, x_2, \ldots, x_r \) are random and have a joint distribution

\[
f(y, x_1, x_2, \ldots, x_r),
\]

which is not necessarily normal so that

S. S. Ganguly is a Professor in the Department of Family Medicine and Public Health.
Email at: ganguly@squ.edu.om
S. S. GANGULY

\[ f(y, x_1, x_2, ..., x_r) = g(y \mid x_1, x_2, ..., x_r) \prod_{i=1}^{r} h(x_i). \]  

(1)

It is assumed herein that the conditional distribution of \( y \) given \( x_1, x_2, ..., x_r \) is normal and is given by

\[
g(y \mid x_1, x_2, ..., x_r) \approx \left\{ \sigma_y^2 (1 - \sum_{i=1}^{r} \rho_{oi}^2) \right\}^{\frac{1}{2}} \times \exp \left\{ -\frac{1}{2\sigma_y^2 (1 - \sum_{i=1}^{r} \rho_{oi}^2)} \left\{ y - \mu_o - \sum_{i=1}^{r} \rho_{oi} \left( \frac{\sigma_o}{\sigma_i} \right) (x_i - \mu_i) \right\}^2 \right\} \]  

(2)

with mean

\[
E(y \mid x) = \mu_o - \sum_{i=1}^{r} \rho_{oi} \left( \frac{\sigma_o}{\sigma_i} \right) (x_i - \mu_i) \]  

(3)

and variance

\[
V(y \mid x) = \sigma_o^2 \left( 1 - \sum_{i=1}^{r} \rho_{oi}^2 \right). \]  

(4)

The marginal density corresponding to the covariate \( x_i \) is assumed to be symmetric about mean of the form:

\[
\frac{1}{\sigma_i} f \left( \frac{x_i - \mu_i}{\sigma_i} \right) \]  

(5)

Here \( \mu_i = E(x_i), \sigma_i^2 = V(x_i) \) and \( \rho_{oi} (i = 1, 2, ..., r) \) is the correlation coefficient between \( y \) and \( x_i \). Relation (2) provides for the measurement of dependency of the response random variable on the random covariates \( x_i \) (\( i=1,2,\ldots,r \)).

The linear relationship may also be written in the form of classical regression model as

\[
E(y \mid x) = \alpha + \sum_{i=1}^{r} \beta_i x_i \]  

(6)
where
\[ \alpha = \mu_o - \sum_{i=1}^{r} \rho_{oi} \left( \frac{\sigma_o}{\sigma_i} \right) \mu_i \] (7)

and
\[ \beta_i = \rho_{oi} \left( \frac{\sigma_o}{\sigma_i} \right), \quad i = 1, 2, \ldots, r \] (8)

are the regression coefficients. It may be noted that \( E(y|x) \) is the best linear predictor of the response variable \( y \) where the population is \( N_r(\mu, \Sigma) \).

In medical epidemiology, one often encounters situations where some (if not all) covariates \( x_i \) have non-normal symmetric distributions. This article is restricted to a situation where the covariates have non-normal symmetric distributions. The objective, therefore, is to estimate the parameters \( (\alpha, \beta)^T \) from \( n \) sample values \( (y_i, x_i), 1 \leq i \leq n \). For this, consider the family of student’s \( t \)-distributions. The method, which has been developed here, is, of course, general and can be used for other families of location-scale distributions of the type (5).

**Likelihood equations**

Suppose that the covariate \( x_i \) (\( i=1, 2, \ldots, r \)) has the symmetric distribution with the density given by
\[ h(x_i) \approx \left( k_i \sigma_i^2 \right)^{\frac{1}{2}} \left\{ 1 + \left( \frac{x_i - \mu_i}{k_i \sigma_i} \right)^2 \right\}^{-p_i}, -\infty < x_i < \infty \] (9)

where \( k_i = 2p_i - 3, \quad p_i \geq 2; \quad E(x_i) = \mu_i \) and \( v(x_i) = \sigma_i^2 \). Assume that \( p_i \) is known. For \( p_i = 5 \), (9) is almost indistinguishable from logistic distribution, because the two distributions are both symmetric and have first four moments common (Pearson, 1963). If the two distributions are plotted, it will be seen that one sits almost on top of the other. It may be noted that
\[ t = \left( \frac{v_i(x_i - \mu_i)}{\sigma_i(k_i)^{\frac{1}{2}}} \right) \]

has Student’s t – distribution with \((2p_i - 1)\) degrees of freedom. For \(1 \leq p_i \leq 2\), \(k\) is equal to 1 in which case \(\sigma\) in (9) is simply a scale parameter.

Given the data matrix \((n > r+1)\) of the form

\[(y_j; x_{j1}, \ldots, x_{ji}, \ldots, x_{jk}), j = 1, 2, \ldots, n\] (10)

where \(y\) is the response variable and the \(x\) terms as explanatory variables or covariates. Then the likelihood function \(L\) based on relation (1) can be written as usual and is given by

\[ L \approx \left\{ \prod_{j=1}^{r} \sigma_{o_i}^2 \left( 1 - \sum_{i=1}^{r} \rho_{oi}^2 \right) \right\}^{-\frac{n}{2}} \]

\[ \times \exp \left[ -\frac{1}{2\sigma_o^2 \left( 1 - \sum_{i=1}^{r} \rho_{oi}^2 \right)} \sum_{j=1}^{n} \left( y_{ij} - \mu_o - \sum_{i=1}^{r} \rho_{oi} \left( \sigma_o \sigma_i \right) \left( x_{ijo} - \mu_i \right) \right)^2 \right] \]

(11)

\[ \times \prod_{j=1}^{n} \prod_{i=1}^{r} \left( 1 + \frac{(x_i - \mu_i)^2}{k_i \sigma_i^2} \right)^{-\frac{p}{2}} \]

where \(x_{ijo}\); \((i = 1, 2, \ldots, r; j = 1, 2, \ldots, n)\) are the order statistics of \(x_i\) observations, and \(y_{ij}\) \((j = 1, \ldots, n)\) are the corresponding concomitant \(y\) observations. The maximum likelihood estimators are the solutions of the likelihood equations, i.e, of the derivatives of \(nL\). These equations are, however, intractable. Solving them by iterative procedures may be problematic, for example, one may encounter multiple roots, slow convergence, or convergence to wrong values (see specifically Barnett, 1966; Lee et al., 1980; Tiku and Suresh, 1992; Vaughan, 1992). Instead the Tikus method of modified likelihood (MML) estimation was employed, which gives explicit estimators and involves replacing intractable terms by linear approximations. Because this method is already well established
and is known to produce estimators which are fully efficient for large \( n \) (Tiku, 1970; Bhattacharyya, 1985) and almost fully efficient for small \( n \) (Tiku et al, 1986; Tiku and Suresh, 1992; Vaughan, 1992, 1994).

**Modified Maximum Likelihood**

Consider the \( i^{th} \) covariate of a random sample of size \( n \) denoted by \( x_{1i}, x_{2i}, \ldots, x_{ni} \) from any location-scale distribution with density given by

\[
\frac{1}{\sigma_i} f \left[ \frac{x_n - \mu_i}{\sigma_i} \right], \quad i = 1, 2, \ldots, r.
\]

For simplicity of notation, suppress the suffix \( i \) and consider \( f \) to be a student \( t \) density. Then the likelihood equations for estimating \( \mu \) and \( \sigma \) corresponding to each covariate are

\[
\frac{\partial (\ln L)}{\partial \mu} = \frac{2p}{k\sigma} \sum_{j=1}^{n} g(z_j) = 0
\]

and

\[
\frac{\partial (\ln L)}{\partial \sigma} = \frac{-n}{\sigma} + \frac{2p}{k\sigma} \sum_{j=1}^{n} z_j g(z_j) = 0
\]

where

\[
z_j = \frac{(x_j - \mu)}{\sigma}
\]

and

\[
g(z_j) = \frac{z_j}{\left\{1 + (1/k)z_j^2 \right\}}.
\]

Equations (12) do not provide explicit solutions. Following Tiku-Suresh (1992); Vaughan and Tiku (2000), the first step is to express these equations in terms of order statistics \( x_{(1)} \leq x_{(2)} \ldots \leq x_{(n)} \). Because complete sums are invariant to ordering
\[ \frac{\partial \ln L}{\partial \mu} = \frac{2p}{k\sigma} \sum_{j=1}^{n} g(z_{(j)}) = 0 \]

and
\[ \frac{\partial \ln L}{\partial \sigma} = \frac{-n}{\sigma} + \frac{2p}{k\sigma} \sum_{j=1}^{n} z_{(j)} g(z_{(j)}) = 0 \] \hspace{1cm} (13)

where
\[ z(j) = \frac{(x_{(j)} - \mu)}{\sigma}, \quad j = 1, 2, \ldots, n. \]

Under appropriate regularity considerations which are very general in nature, \( g(z_{(j)}) \) can be replaced by linear approximations given by the first two terms of Taylor series expansions (Tiku, 1967, 1968; Tiku and Suresh, 1992; Tiku and Kambo, 1992, Vaughan, 1992; Vaughan and Tiku, 2000), so that
\[ g\{z_{(j)}\} \approx g\{t_{(j)}\} + \left[ z_{(j)} - t_{(j)} \right] \left( \frac{d}{dz} g(z) \right)_{z=t_{(j)}} = \alpha_j + \beta_j z_{(j)}, \quad j = 1, 2, \ldots, n \] \hspace{1cm} (14)

where
\[ t_{(j)} = E\{z_{(j)}\}. \]

Thus, the modified equations are obtained, i.e.
\[ \frac{\partial \ln L}{\partial \mu} \approx \frac{\partial \ln L^*}{\partial \mu} = \frac{2p}{k\sigma} \sum_{j=1}^{n} \{ \alpha_j + \beta_j z_{(j)} \} = 0 \]

and
\[ \frac{\partial \ln L}{\partial \sigma} \approx \frac{\partial \ln L^*}{\partial \sigma} = \frac{-n}{\sigma} + \frac{2p}{k\sigma} \sum_{j=1}^{n} z_{(j)} \{ \alpha_j + \beta_j z_{(j)} \} = 0 \] \hspace{1cm} (15)

Equations (15) have explicit solutions, which are called modified maximum likelihood (MML) estimators. Note that the ML and MML estimators are asymptotically equivalent.

For distribution \( (p \geq 2, k = 2p - 3) \)
ROBUST REGRESSION ANALYSIS FOR NON-NORMAL SITUATIONS

\[ h(x_j) \propto (k\sigma)^{-\frac{1}{2}} \left\{ 1 + \frac{x_{(j)} - \mu}{k\sigma^2} \right\}^{-p}, \ -\infty \leq x_j \leq \infty \] (16)

This method gives the following MML estimators (see Tiku and Suresh, 1992; Tiku and Kambo, 1992; Vaughan, 1992; Vaughan and Tiku, 2000; Tiku et al, 2008)

\[ \hat{\mu} = \frac{1}{m} \sum_{j=1}^{n} \beta_j x_{(j)} \quad (m = \sum_{j=1}^{n} \beta_j) \] (17)

and

\[ \hat{\sigma} = \left\{ B + (B^2 + 4nc)^{\frac{1}{2}} \right\} \frac{1}{2\{n(n-1)\}} \] (18)

where

\[ B = \frac{2p}{k} \sum_{j=1}^{n} \alpha_j x_{(j)} \text{ and } C = \frac{2p}{k} \left\{ \sum_{j=1}^{n} \beta_j y_{(j)}^2 - m \mu^2 \right\} \] (19)

The coefficients \( \alpha_j \) and \( \beta_j \) are obtained from the equations

\[ \alpha_j = \left\{ \frac{2}{k} \frac{t^3_{(j)}}{1 + \frac{1}{k} t^2_{(j)}} \right\} \text{ and } \beta_j = \frac{1 - (1/k) t^2_{(j)}}{\left\{1 + (1/k) t^2_{(j)}\right\}^2}, \ j = 1, 2, ..., n \] (20)

For \( p = \infty \) (i.e. for normal distribution), \( \alpha_j = 0 \) and \( \beta_j = 1 \), because \( k = 2p - 3 \). Note that \( \alpha_j = -\alpha_{n-j+1}, \beta_j = \beta_{n-j+1} \ (1 \leq j \leq n) \) and \( \sum_{j=1}^{n} \alpha_j = 0 \). Tables of the value of \( t_{(j)} \) are available for \( p = 2(.5) 10 \) and \( n \leq 20 \) (Tiku and Kumra, 1985). For \( n > 20 \), \( t_{(j)} \) are obtained from the equation
In evaluating (21), it should be noted that \( \{(k / \nu)z\}^{1/2} \) has student’s t-distribution with \( \nu = 2p - 1 \) degrees of freedom.

It may be of interest to note that in deriving the estimators \( \mu \) and \( \sigma \) given by the equations (17)-(20), the method of MML estimation for \( p < \infty \) automatically gives small weights to extreme order statistics close to the center. It is precisely due to this fact these estimators are robust to reasonable departures from the true value of \( p \) in (16). In most applications, therefore, it is not very important to pinpoint the true value of \( p \) and use it in all derivatives. Any reasonable value of \( p \) gives almost optimal results.

A Q-Q plot can be employed to give a reasonable value closure (if not exactly) the true value of \( p \) corresponding to covariate \( x \) (Tiku et al, 1986, p.277). The order statistic \( x_{(j)} \) is plotted against the values \( t_{(j)} = E(z_{(j)}), z_j = (x_j - \mu) / \sigma, j = 1,2,...,n \), under the assumed model, i.e. for a particular value of \( p \) in (16). If the plot gives a straight line (or nearly so), the model is taken to be valid for the MML estimation.

Following the above procedure, the parameters \( \mu_i \) and \( \sigma_i (i = 1,2,...,r) \) are estimated. In order to estimate the remaining parameters viz., \( \mu_o, \sigma_o, \rho_{oi} (i = 1,2,...,r) \), the likelihood function (11) is considered. Because
\[
\frac{\partial \ln L}{\partial \mu_i} \quad \text{and} \quad \frac{\partial \ln L}{\partial \sigma_i} \quad (i=1,2,...,r)
\]
are expressed in terms of \( g \{z_{(i)}\} \), the likelihood equations
\[
\frac{\partial \ln L}{\partial \mu_i} = 0, \quad \frac{\partial \ln L}{\partial \sigma_i} = 0 \quad (i = 1,2,...,r)
\]
and
\[
\frac{\partial \ln L}{\partial \rho_{oi}} = 0 \quad (i = 1,2,...,r)
\]
have no explicit solutions. The modified likelihood equations are
\[
\frac{\partial \ln L^*}{\partial \mu_i} = 0, \quad \frac{\partial \ln L^*}{\partial \sigma_i} = 0 \quad (i = 0,1,...,r)
\]
and
\[
\frac{\partial \ln L^*}{\partial \rho_{oi}} = 0 \quad (i = 1,2,...,r),
\]
and are obtained by replacing \( g \{z_{(i)}\} \) with the linear approximations given by (14). The solutions of these equations are the following MML estimators:
\[
\hat{\mu}_o = y - \sum_{i=1}^{r} \rho_{oi} \left( \frac{\hat{\sigma}_o}{\sigma_i} \right) (x_i - \hat{\mu}_i)
\]
(22)
ROBUST REGRESSION ANALYSIS FOR NON-NORMAL SITUATIONS

\[ \hat{\sigma}_o = \left[ S_o^2 + \sum_{i=1}^{r} \left( \frac{S_o^2}{S_i} \left( \frac{\sigma_i^2}{S_i} - 1 \right) \right) \right]^{1/2} \]  

(23)

\[ \hat{\rho}_{oi} = \frac{S_{oi}^2}{S_i S_o} \], \quad i = 1, 2, \ldots, r \]  

(24)

Here,

\[ n \bar{y} = \sum_{j=1}^{n} y_{[j]} = \sum_{j=1}^{n} y_j, \quad n \bar{x} = \sum_{j=1}^{n} x_{(j)} = \sum_{j=1}^{n} x_j \]  

(25)

\[ (n-1)S_o^2 = \sum_{j=1}^{n} \left[ y_{[j]} - \bar{y} \right]^2 = \sum_{j=1}^{n} \left( y_j - \bar{y} \right)^2 \]  

(26)

\[ (n-1)S_i^2 = \sum_{j=1}^{n} \left[ x_{(j)} - \bar{x}_i \right]^2 = \sum_{j=1}^{n} \left( x_j - \bar{x}_i \right)^2 \]  

(27)

and

\[ (n-1)S_{oi}^2 = \sum_{j=1}^{n} \left[ y_{[j]} - \bar{y} \right] \left[ x_{(j)} - \bar{x}_i \right] \]  

(28)

\[ = \sum_{j=1}^{n} \left[ y_j - \bar{y} \right] \left[ x_{ij} - \bar{x}_i \right], \quad i = 1, 2, \ldots, r. \]

Relation (22) provides for the measurement of dependency of the response random variable on the random covariates \( x_i \) \( (i = 1, 2, \ldots, r) \). The linear relationship is also represented in the form of classical model (6).

The asymptotic variances and covariances of the estimators \( \hat{\mu}_o, \hat{\mu}_i, \hat{\sigma}_o, \hat{\sigma}_i \) and \( \hat{\rho}_{oi} \) \( (i = 1, 2, \ldots, r) \) are obtained with the use of the second partial derivatives of the likelihood function (11). The matrix formed by the negative of the expected values of the second partial derivatives gives the information matrix, which may be expressed as the partitioned matrix.
\[ V = \begin{pmatrix} V_1 & O \\ O & V_2 \end{pmatrix} \]  

where the matrix is of the order \((3r+2) \times (3r+2)\) and

\[ V_1 = E \left[ -\frac{\partial^2 \ell (nL' \hat{\theta})}{\partial \mu_i \partial \mu_i} \right] \]

of order \((r+1) \times (r+1)\) and

\[ V_2 = E \left[ -\frac{\partial^2 \ell (nL' \hat{\theta})}{\partial \theta_i \partial \theta_i} \right], \quad i, i' = 1, 2, \ldots, r \]

of order \((2r+1) \times (2r+1)\) with \((\hat{\theta}_i = \sigma_1, \theta_2 = \sigma_2, \ldots, \theta_{2r+1} = \rho_{ab})\).

The inverse of \(V_1\) and \(V_2\) matrices provides the elements of the precision and covariance structure of the estimated coefficients.

The estimated values of the parameters obtained above are used in relation (7) and (8) which give the estimated values of the regression coefficients \(\hat{\alpha}\) and \(\hat{\beta}_i\) \((i = 1, 2, \ldots, r)\) of the model (6). The asymptotic covariance structure of the estimated regression coefficients \(\hat{\alpha}\) and \(\hat{\beta}_i\) \((i = 1, 2, \ldots, r)\) are obtained using delta method (Serfling, 1980) as:

Let \(g(\alpha, \beta)\) and \(\hat{\theta} = (\mu, \sigma, \rho)\), then

\[ \begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} \approx N_{r+1} \left[ \begin{pmatrix} \alpha \\ \beta \end{pmatrix}, \quad n^{-1} G^T \hat{\Sigma} G \right], \]

where

\[ G = \left[ \frac{\partial g}{\partial \hat{\theta}} \right]_{\mu, \ldots, \rho_{ab}} \]

of order \((3r+2) \times (r+1)\) and
ROBUST REGRESSION ANALYSIS FOR NON-NORMAL SITUATIONS

\[
\hat{\Sigma} = \begin{pmatrix}
V_1 & O \\
O & V_2^{-1}
\end{pmatrix}
\]

of order \((3r+2) \times (3r+2)\). Note that when \(p = \infty\) the distribution (16) reduces to the ideal normal distribution in which case \(\hat{\mu} = \bar{x}\) (sample mean) and \(\hat{\sigma}^2 = s^2\) (sample variance), \(\bar{x}\) and \(s^2\) being optimal under the assumption of normality.

Examples

Example 1

Consider the part of the data set pertaining to 20 male insulin-dependent diabetic patients as provided in Dobson (1990, p. 69), which is reproduced in Table 1.

<table>
<thead>
<tr>
<th>(y = \text{Carb. (gm)})</th>
<th>(x_1 = \text{Age (yrs)})</th>
<th>(x_2 = \text{Wgt (kg)})</th>
</tr>
</thead>
<tbody>
<tr>
<td>33</td>
<td>33</td>
<td>100</td>
</tr>
<tr>
<td>40</td>
<td>47</td>
<td>92</td>
</tr>
<tr>
<td>37</td>
<td>49</td>
<td>135</td>
</tr>
<tr>
<td>27</td>
<td>35</td>
<td>144</td>
</tr>
<tr>
<td>30</td>
<td>46</td>
<td>140</td>
</tr>
<tr>
<td>43</td>
<td>52</td>
<td>101</td>
</tr>
<tr>
<td>34</td>
<td>62</td>
<td>95</td>
</tr>
<tr>
<td>48</td>
<td>23</td>
<td>101</td>
</tr>
<tr>
<td>30</td>
<td>32</td>
<td>98</td>
</tr>
<tr>
<td>38</td>
<td>42</td>
<td>105</td>
</tr>
</tbody>
</table>

In this sample, the goal is to establish the relationship between the response variable \(y\) (amount of carbohydrate) and the two covariates \(x_1\) (age) and \(x_2\) (body weight, relative to “ideal” weight for height) using the linear regression model (6) which takes the form

\[
E(y|x_1, x_2) = \alpha + \beta_1 x_1 + \beta_2 x_2
\]

(31)

Here, it is assumed that, in relation (1), the conditional distribution of the response random variable \(y\) is normal; however, the covariates follow
independently non-normal symmetric distribution. The model (31) is fitted using above described modified maximum likelihood method.

First obtain the values of $p_1$ and $p_2$ corresponding to the two covariates $x_1$ and $x_2$ using Q-Q plots, where the order statistics $x_{1(j)}$ and $x_{2(j)}$ were plotted separately against $t_{1(j)}$ and $t_{2(j)}$ respectively, $j = 1, \ldots, n$ for different values of $p$ as given in Tiku and Kumra (1985). The values of $p_1 = 5$ and $p_2 = 7$ provided an approximate straight line patterns which determined the appropriate types of densities in (16). Once $p_1$ and $p_2$ are known, then using the equations (17)-(20), the MML estimates of the parameters $\mu_1, \sigma_1$ and $\mu_2, \sigma_2$ are obtained. Using these values in equations (22)-(28) the rest of the parameters $\mu_o, \sigma_o, \rho_{o1}$ and $\rho_{o2}$ are estimated. Solutions of the information matrix (29) provided the elements of the precision and covariance structure of the estimated parameters. The estimated values and their standard errors are presented in Table 2.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_0$</td>
<td>37.732</td>
<td>1.848</td>
</tr>
<tr>
<td>$\mu_1$</td>
<td>46.437</td>
<td>3.008</td>
</tr>
<tr>
<td>$\mu_2$</td>
<td>109.936</td>
<td>3.776</td>
</tr>
<tr>
<td>$\sigma_0$</td>
<td>7.635</td>
<td>1.411</td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>13.989</td>
<td>1.789</td>
</tr>
<tr>
<td>$\sigma_2$</td>
<td>17.265</td>
<td>2.351</td>
</tr>
<tr>
<td>$\rho_{o1}$</td>
<td>-0.064</td>
<td>0.228</td>
</tr>
<tr>
<td>$\rho_{o2}$</td>
<td>-0.420</td>
<td>0.243</td>
</tr>
</tbody>
</table>

Using the estimated values in Table 2 in relation (7) and (8), obtain MML estimates of the regression parameters $\alpha, \beta_1$ and $\beta_2$. Use of delta method as described in (30) provided the asymptotic standard errors; also these parameters based on usual maximum likelihood method were estimated. The results, obtained under the two methods are summarized in Table 3.

The analysis in Table 3 reveals that the MML estimates of the regression parameters for the data set in Table 1 are very close to the values obtained using maximum likelihood method, as expected. Moreover, the two methods gave approximately the same results for the Wald statistics $W$, which permits to test the
null hypothesis $H_0: \beta_1 = 0$ and $\beta_2 = 0$. For large $n$, the null distribution of $W$ is referred to a standard normal distribution.

**Example 2**

Consider another data set from Murray (1937), reproduced in El-Saidi (1995, p. 214) as shown in Table 4. The data provides 11 observations on the number of male flies died after twenty minutes exposure to pyrethrum at various concentrations.

The main objective is to describe the probability of success $p_j$ as a function of dose $x_j$. In literature, such type of analysis are carried out usually considering either probit or logit models (Cox, 1970). However, the logit model is preferred to a probit model due to two primary reasons (Hosmer and Lameshow, 1989): from mathematical point of view, it is an easily used function, and it leads to itself to a biological meaningful interpretation.

**Table 4.** Mortality of male flies after twenty minutes exposure to pyrethrum

<table>
<thead>
<tr>
<th>Concentration (log$10$)</th>
<th>Number of flies</th>
<th>Proportions Died</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Exposed</td>
<td>Died</td>
</tr>
<tr>
<td>1.6020</td>
<td>462</td>
<td>109</td>
</tr>
<tr>
<td>1.7782</td>
<td>500</td>
<td>199</td>
</tr>
<tr>
<td>1.9031</td>
<td>467</td>
<td>298</td>
</tr>
<tr>
<td>2.0000</td>
<td>515</td>
<td>370</td>
</tr>
<tr>
<td>2.0792</td>
<td>561</td>
<td>459</td>
</tr>
<tr>
<td>2.1461</td>
<td>469</td>
<td>400</td>
</tr>
<tr>
<td>2.2041</td>
<td>550</td>
<td>495</td>
</tr>
<tr>
<td>2.2553</td>
<td>542</td>
<td>499</td>
</tr>
<tr>
<td>2.3010</td>
<td>479</td>
<td>450</td>
</tr>
<tr>
<td>2.3979</td>
<td>497</td>
<td>476</td>
</tr>
<tr>
<td>2.4771</td>
<td>453</td>
<td>442</td>
</tr>
</tbody>
</table>

The logit model is a family of Generalized Linear Models (GLMs) with link function $g(p_j)$ as $\ln\left(\frac{p_j}{1-p_j}\right)$ (Nelder and Wedderburn, 1972; McCullagh and Nelder, 1989). The link function $g(p_j)$ is continuous and maps the $[0,1]$ range of probabilities onto $(-\infty, \infty)$ and is represented by
so that

\[ p_j = \frac{\exp(\alpha + \beta x_j)}{1 + \exp(\alpha + \beta x_j)} , j = 1, 2, \ldots, n \]  \hspace{1cm} (33)

The relation (33) is known as binary logistic model with probability of success \( p_j \), this belongs to the standardized logistic distribution which is symmetric in nature (Rao and Toutenburg, 1995, p. 263).

In order to estimate the unknown parameters \( \alpha \) and \( \beta \) in (32), usually ML method is used. The technique involves the solution of the likelihood equations, which have no explicit solutions and have to be solved by interactive procedures. Solving these equations is, therefore, tedious and time consuming. Therefore, these parameters are estimated using MML method.

For this, consider the link function i.e. log odds as a response variable and \( x_j \) as a covariate. First estimate \( \mu_1 \) and \( \sigma_1 \) for \( p=5 \) in distribution (16). Using these values in equations (22)-(28), the rest of the parameters \( \mu_o, \sigma_o \) and \( \rho_{o1} \) involved in the likelihood function (11) were obtained. The estimated values of the variances and co-variances were obtained using these values in second partial derivatives of the likelihood function (11) and solving for the inverse of the information matrix (29). The estimated values of the parameters and their standard errors involved in the likelihood function (11) with \( p =5 \) for the data set in Table 4 are shown in Table 5.

Using these estimated values of the parameters in relation (7) and (8), obtain the MML estimates of the parameters \( \hat{\alpha} \) and \( \hat{\beta} \) of the logistic model (33). The use of delta method (30) gave the asymptotic variances of \( \hat{\alpha} \) and \( \hat{\beta} \). The ML estimates of these parameters and their variances under the logit model (32) were also obtained using iterative procedures viz; Newton-Raphson method (Cox, 1970, Chapter 2). The results obtained under the two procedures are summarized in Table 6.

These analyses also reveal that the MML estimates of the regression parameters \( \alpha \) and \( \beta \) for the data set in Table 4 are very close to the values obtained using maximum likelihood method, as expected.
Table 5. MML estimates of the parameters and their standard errors for the data set in Table 4

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_0$</td>
<td>1.642</td>
<td>0.459</td>
</tr>
<tr>
<td>$\mu_1$</td>
<td>2.115</td>
<td>0.082</td>
</tr>
<tr>
<td>$\sigma_0$</td>
<td>1.593</td>
<td>0.198</td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>0.284</td>
<td>0.035</td>
</tr>
<tr>
<td>$\rho_{01}$</td>
<td>0.999</td>
<td>0.003</td>
</tr>
</tbody>
</table>

Table 6. MML and ML estimates of the parameters and their standard errors for logit model (32)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>MML</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant ($\alpha$)</td>
<td>-10.219</td>
<td>0.186</td>
</tr>
<tr>
<td>Coefficient ($\beta$)</td>
<td>5.608</td>
<td>0.087</td>
</tr>
<tr>
<td>ML</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant ($\alpha$)</td>
<td>-10.329</td>
<td>0.343</td>
</tr>
<tr>
<td>Coefficient ($\beta$)</td>
<td>5.661</td>
<td>0.172</td>
</tr>
</tbody>
</table>

This study used Tiku’s modified maximum likelihood method for carrying out regression analysis when the underlying distributions of the data set have non-normal symmetric distributions. The method yields estimators which are explicit functions of sample observations and are numerically very close to the maximum likelihood estimators and equally efficient.

References


