A RANDOM-EFFECTS LOG-LINEAR MODEL WITH POISSON DISTRIBUTIONS

Authors: Maria Alexandra Seco
– Department of Mathematics of ESTG, Instituto Politécnico de Leiria, Portugal (aseco@estg.ipleiria.pt)

António St.Aubyn
– Department of Mathematics of ISA, Universidade Técnica de Lisboa, Portugal (staubyn@isa.utl.pt)

Received: October 2002 Revised: April 2003 Accepted: June 2003

Abstract:

- In several applications data are grouped and there are within-group correlations. With continuous data, there are several available models that are often used; with counting data, the Poisson distribution is the natural choice. In this paper a mixed log-linear model based on a Poisson–Poisson conditional distribution is presented. The initial model is a conditional model for the mean of the response variable, and the marginal model is formed thereafter. Random effects with Poisson distribution are introduced and a variance-covariance matrix for the response vector is formed embodying the covariance structure induced by the grouping of the data.

Key-Words:

- *log-linear models; grouped data; random effects; mixed models; overdispersion; iterative reweighted generalized least squares.*

AMS Subject Classification:

- 62J02, 62J12, 62J99, 62P12.

1. INTRODUCTION

In many applications in biology, agriculture, engineering and economics, for instance, grouped data reveal within-group correlation. For continuous data there are several available models which are used. These include Variance Component Models and Mixed Models (Laird and Ware [2], Pinheiro and Bates [6]) which embody fixed and random effects. Both models are based on the Multivariate Normal distribution, which has friendly properties, as the marginal and conditional distributions are still Normal.

Goldstein [1] gives several examples where ignoring the group structure can lead to imprecise estimates, confidence intervals and significant tests. He alerts that grouped data should be modelled respecting its particular structure.

A mixed log-linear model based on the Poisson–Poisson hierarchical distribution will be presented for grouped count data. The initial model is a conditional model for the mean of Y, and the marginal model is derived afterwards. It will be shown that building the model this way and introducing random Poisson effects, is a means of introducing overdispersion in a pseudo-Poisson model (overdispersion is said to exist when $\mathrm{var}(Y) = \phi\, E(Y),\ \phi > 1$). Moreover, the variance-covariance matrix is built for the response vector $\mathbf{Y}$, which embodies the covariance structure induced by the grouping of the data.

Several authors (McCulloch and Searle [5], Vonesh and Chinchilli [7]) have made references to some mixed models based on Poisson–Gamma or Bernoulli–Beta distributions as they are conjugate families. Starting from a model where $Y_{ij}|b_i$ follows a Poisson law and b_i a Gamma one, and as the $Y_{ij}|b_i$ are conditionally independent, the derived density function for $\mathbf{Y}_i$, a density product, is computationally unfriendly. In this paper a practical and simpler approach is proposed, that starts from a Poisson–Poisson model and uses the marginal moments of the response variable. The parameters are then estimated, with the iterative, non-linear, generalized least squares method.

In this presentation, attention is given to the simplest case of a single random effect. This is not as restrictive as it seems because, as was referred above, it portrays a situation of overdispersion with within-group correlation.

2. THE LOG-LINEAR CONDITIONAL MODEL

Consider M groups, with n_i observations per group (counts), where a within-group correlation structure is expected. Define the mixed log-linear model

$$\log\left[E(\mathbf{Y}_i|b_i)\right] \;=\; \mathbf{X}_i\boldsymbol{\beta} + \mathbf{1}_{n_i} b_i\ , \qquad i = 1,...,M,\quad j = 1,...,n_i\ . \tag{2.1}$$

Here $\mathbf{Y}_i = [Y_{i1} \, ... \, Y_{in_i}]^T$ is a random vector $n_i{\times}1$, b_i is a random variable $(1{\times}1)$, $\mathbf{X}_i$ is a known model matrix of order $n_i{\times}p$, $\boldsymbol{\beta}$ is a $p{\times}1$ vector of unknown fixed parameters and $\mathbf{1}_{n_i}$ is a vector $n_i{\times}1$ of ones. $\mathbf{Y}_i$ and b_i are independent for different i's.

Consider that each $Y_{ij}|b_i$ is a random variable conditionally independent for different j's following the Poisson law

$$Y_{ij}|b_i \sim P\Big(\exp\big\{\mathbf{x}_j^T\boldsymbol{\beta} + b_i\big\}\Big) , \qquad i = 1, ..., M, \quad j = 1, ..., n_i ,$$

where $\mathbf{x}_j^T$ is row j of the model matrix $\mathbf{X}_i$ and $\boldsymbol{\beta}$ is the same as before. Let

$$\begin{aligned} &b_i \sim P(\theta_i) , \\ &\theta_i > 0 , \end{aligned}$$

independent for different i's.

Hence $E(b_i) = \text{var}(b_i) = \theta_i, \ i = 1, ..., M$.

Note that $\mathbf{Y}$, the vector of all the random variables, is an $N{\times}1$ vector which is partioned as M components $\mathbf{Y}_i$, each of which is a random n_i-vector, $i = 1..., M$,

$$\mathbf{Y} = \begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \vdots \\ \mathbf{Y}_M \end{bmatrix} = \begin{bmatrix} Y_{11} \\ Y_{12} \\ \vdots \\ Y_{1n_1} \\ Y_{21} \\ \vdots \\ Y_{Mn_M} \end{bmatrix} .$$

N is the total number of observations, $N = \sum_{i=1}^{M} n_i$. Note that $\text{cov}(Y_{ij}, Y_{ik}) \neq 0$, $j \neq k$, i.e., the Y_{ij} for $j = 1, .., n_i$, are not independent as they represent the same group, but they are independent for different i's (groups). Each b_i random variable is introduced to portray the situation of within group correlation for group i, $i = 1, ..., M$.

3. THE MARGINAL MODEL FOR Y

The parameter estimates are computed from a model based on the marginal moments of $\mathbf{Y}$. The mean value, variance and covariance of the $\mathbf{Y}$ marginals are then computed.

Let Y_{ij} be the variable that corresponds to the j-th observation in group i, $i = 1, ..., M, \ j = 1, ..., n_i$. As it is assumed that $Y_{ij}|b_i \sim P\left(\exp\left\{\mathbf{x}_j^T\boldsymbol{\beta} + b_i\right\}\right)$

and $b_i \sim P(\theta_i)$,

$$\begin{aligned} E(Y_{ij}) &= E_{b_i}\Big[E(Y_{ij}|b_i)\Big] \\ &= E\Big(\exp\{\mathbf{x}_j^T\boldsymbol{\beta} + b_i\}\Big) \\ &= \exp\{\mathbf{x}_j^T\boldsymbol{\beta}\}M_{b_i}(1) \ , \end{aligned}$$

where $M_{b_i}(\cdot)$ is the b_i moment generating function. Then

$$\begin{aligned} E(Y_{ij}) &= \exp\{\mathbf{x}_j^T\boldsymbol{\beta}\}\ \exp\{(e-1)\,\theta_i\} \\ &= \exp\Big\{\mathbf{x}_j^T\boldsymbol{\beta} + (e-1)\,\theta_i\Big\} \ , \end{aligned}$$

where e is the Neper number, and

$$\log\left[E(Y_{ij})\right] = \mathbf{x}_j^T\boldsymbol{\beta} + (e-1)\,\theta_i \ .$$

Note the offset, $(e-1)\,\theta_i$, that comes out in the marginal expected value of Y_{ij}, derived from the introduction of the random effect b_i in the conditional model.

For the Y_{ij} variance,

$$\begin{aligned} \mathrm{var}(Y_{ij}) &= \mathrm{var}\Big[E(Y_{ij}|b_i)\Big] + E\Big[\mathrm{var}(Y_{ij}|b_i)\Big] \\ &= \mathrm{var}\Big(\exp\{\mathbf{x}_j^T\boldsymbol{\beta} + b_i\}\Big) + E\Big(\exp\{\mathbf{x}_j^T\boldsymbol{\beta} + b_i\}\Big) \\ &= E\Big(\exp\Big\{2(\mathbf{x}_j^T\boldsymbol{\beta} + b_i)\Big\}\Big) - \Big[E\Big(\exp\Big\{\mathbf{x}_j^T\boldsymbol{\beta} + b_i\Big\}\Big)\Big]^2 \\ &\quad + E\Big(\exp\Big\{\mathbf{x}_j^T\boldsymbol{\beta} + b_i\Big\}\Big) \\ &= \exp\{\mathbf{x}_j^T\boldsymbol{\beta}\}\ \Big[\exp\{\mathbf{x}_j^T\boldsymbol{\beta}\}M_{b_i}(2) - \exp\{\mathbf{x}_j^T\boldsymbol{\beta}\}(M_{b_i}(1))^2 + M_{b_i}(1)\Big] \\ &= E(Y_{ij})\left[\exp\{\mathbf{x}_j^T\boldsymbol{\beta}\}\frac{M_{b_i}(2)}{M_{b_i}(1)} - \exp\{\mathbf{x}_j^T\boldsymbol{\beta}\}M_{b_i}(1) + 1\right] \ . \end{aligned}$$

It is known that the distribution of Y_{ij} is not Poisson, but it may be called pseudo-Poisson with overdispersion. Note that

$$\mathrm{var}(Y_{ij}) = \varphi\, E(Y_{ij}) \ ,$$

where the contribution of b_i for the "overdispersion component" is highlighted,

$$\varphi = \exp\{\mathbf{x}_j^T\boldsymbol{\beta}\}\frac{M_{b_i}(2)}{M_{b_i}(1)} - \exp\{\mathbf{x}_j^T\boldsymbol{\beta}\}M_{b_i}(1) + 1 \ .$$

Finally,

$$\begin{aligned}
\operatorname{var}(Y_{ij}) &= \exp\{\mathbf{x}_j^T\boldsymbol{\beta}\}\exp\{(e-1)\,\theta_i\}\times \\
&\quad\times\left[\exp\{\mathbf{x}_j^T\boldsymbol{\beta}\}\,\frac{\exp\{(e^2-1)\,\theta_i\}}{\exp\{(e-1)\,\theta_i\}} - \exp\{\mathbf{x}_j^T\boldsymbol{\beta}\}\exp\{(e-1)\,\theta_i\} + 1\right] \\
&= \exp\{2\,\mathbf{x}_j^T\boldsymbol{\beta}\}\left[\exp\{(e^2-1)\,\theta_i\} - \exp\{2\,(e-1)\,\theta_i\}\right] \\
&\quad + \exp\{\mathbf{x}_j^T\boldsymbol{\beta}\}\exp\{(e-1)\,\theta_i\} \\
&= C(\theta_i)\exp\{2\,\mathbf{x}_j^T\boldsymbol{\beta}\} + K(\theta_i)\exp\{\mathbf{x}_j^T\boldsymbol{\beta}\}\ ,
\end{aligned}$$

where

$$C(\theta_i) = \exp\{(e^2-1)\,\theta_i\} - \exp\{2(e-1)\,\theta_i\}\ ,$$

and

$$K(\theta_i) = \exp\{(e-1)\,\theta_i\}\ . \tag{3.1}$$

For the covariance, with $j \neq k$, and for the i group,

$$\begin{aligned}
\operatorname{cov}(Y_{ij}, Y_{ik}) &= \operatorname{cov}\Big[E(Y_{ij}|b_i), E(Y_{ik}|b_i)\Big] + E\Big[\operatorname{cov}(Y_{ij}, Y_{ik}|b_i)\Big] \\
&= \operatorname{cov}\Big[E(Y_{ij}|b_i), E(Y_{ik}|b_i)\Big] + E(0) \\
&= \exp\{\mathbf{x}_j^T\boldsymbol{\beta} + \mathbf{x}_k^T\boldsymbol{\beta}\}\operatorname{var}[\exp\{b_i\}] \\
&= \exp\{\mathbf{x}_j^T\boldsymbol{\beta} + \mathbf{x}_k^T\boldsymbol{\beta}\}\left[M_{b_i}(2) - (M_{b_i}(1))^2\right] \\
&= \exp\{\mathbf{x}_j^T\boldsymbol{\beta} + \mathbf{x}_k^T\boldsymbol{\beta}\}\left[\exp\{(e^2-1)\,\theta_i\} - \exp\{2(e-1)\,\theta_i\}\right] \\
&= C(\theta_i)\,\exp\{\mathbf{x}_j^T\boldsymbol{\beta} + \mathbf{x}_k^T\boldsymbol{\beta}\}\ .
\end{aligned}$$

3.1. Parameter estimation

The parameter estimates are obtain minimizing

$$\sum_{i=1}^{M}\Big(\mathbf{y}_i - K(\theta_i)\exp\{\mathbf{X}_i\,\boldsymbol{\beta}\}\Big)^T\,\mathbf{V}_i^{-1}\Big(\mathbf{y}_i - K(\theta_i)\exp\{\mathbf{X}_i\boldsymbol{\beta}\}\Big) \tag{3.2}$$

where $\mathbf{y}_i$ is a n_i-dimension vector of responses and $K(\theta_i) = \exp\{(e-1)\,\theta_i\}$, $i = 1, ..., M$. Matrix $\mathbf{V}_i$, the variance-covariance matrix of $\mathbf{Y}_i$, is symmetric of order $n_i \times n_i$, with generic element υ_{jk}:

$$\begin{aligned}
&\mathbf{V}_i = [\upsilon_{jk}]_{j,k=1,\ldots,n_i}\ , \qquad i = 1, ..., M\ , \\
&\upsilon_{jj} = C(\theta_i)\exp\{2\,\mathbf{x}_j^T\boldsymbol{\beta}\} + K(\theta_i)\exp\{\mathbf{x}_j^T\boldsymbol{\beta}\}\ , \\
&\upsilon_{jk} = C(\theta_i)\exp\{\mathbf{x}_j^T\boldsymbol{\beta} + \mathbf{x}_k^T\boldsymbol{\beta}\}\ , \qquad j \neq k\ .
\end{aligned}$$

As $\mathbf{V}_i$ depends on $\boldsymbol{\beta}$ and θ_i it becomes necessary to apply an iterative method. It is possible to apply the IRGLS — Iterative Reweighted Generalized Least Squares method. This is an improvement of the Estimated Generalized Least Squares (EGLS) procedure which iterates using updated values of $\mathbf{V}_i(\hat{\boldsymbol{\beta}}, \hat{\theta}_i)$ to wash out any inefficiency associated with the initial estimates of $\boldsymbol{\beta}$ and θ_i. At each iteration $\mathbf{V}_i$ is updated using current estimates of the parameters. IRGLS may be applied to small or moderate samples (Vonesh and Chinchilli [7]).

Let $\boldsymbol{\theta} = (\theta_1, ..., \theta_M)$ and $\boldsymbol{\tau} = (\boldsymbol{\beta}, \boldsymbol{\theta})$. IRGLS corresponds to solving a set of generalized estimating equations (Liang and Zeger [3]):

$$\mathbf{U}(\boldsymbol{\tau}) = \sum_{i=1}^{M} \mathbf{U}_i(\boldsymbol{\beta}, \theta_i) = \mathbf{0} \ ,$$

or

$$\sum_{i=1}^{M} \left\{ \mathbf{D}_i^T(\boldsymbol{\beta}, \theta_i)\, \mathbf{V}_i^{-1}(\boldsymbol{\beta}, \theta_i) \Big(\mathbf{y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}, \theta_i) \Big) \right\} \;=\; \mathbf{0} \ , \tag{3.3}$$

where $\mathbf{D}_i(\boldsymbol{\beta}, \theta_i) = \dfrac{\partial \boldsymbol{\mu}_i(\boldsymbol{\beta}, \theta_i)}{\partial (\boldsymbol{\beta}, \theta_i)^T}$ and $\boldsymbol{\mu}_i = E(\mathbf{Y}_i)$. A solution to (3.3) can be obtained using the Gauss–Newton algorithm whereby estimates of $\boldsymbol{\tau}$ are updated as

$$\hat{\boldsymbol{\tau}}^{(t+1)} \;=\; \hat{\boldsymbol{\tau}}^{(t)} + \Omega(\hat{\boldsymbol{\tau}}^{(t)})\, \mathbf{U}(\hat{\boldsymbol{\tau}}^{(t)}) \ ,$$

with

$$\Omega(\hat{\boldsymbol{\tau}}^{(t)}) \;=\; \left[\sum_{i=1}^{M} \mathbf{D}_i^T(\hat{\boldsymbol{\beta}}^{(t)}, \hat{\theta}_i^{(t)})\ \mathbf{V}_i^{-1}(\hat{\boldsymbol{\beta}}^{(t)}, \hat{\theta}_i^{(t)})\ \mathbf{D}_i(\hat{\boldsymbol{\beta}}^{(t)}, \hat{\theta}_i^{(t)}) \right]^{-1} .$$

3.2. Inference and asymptotic properties

It is known (Vonesh and Chinchilli [7]) that the $\boldsymbol{\tau}$ IRGLS estimator, under regularity conditions that are usually satisfied, is asymptotically strongly consistent and has a Normal asymptotic distribution with mean zero and variance matrix given by:

$$\Omega(\hat{\boldsymbol{\tau}}) \;=\; \mathrm{var}(\hat{\boldsymbol{\tau}}) \;=\; \left[\sum_{i=1}^{M} \mathbf{D}_i^T(\boldsymbol{\beta}, \theta_i)\, \mathbf{V}_i^{-1}(\boldsymbol{\beta}, \theta_i)\, \mathbf{D}_i(\boldsymbol{\beta}, \theta_i) \right]^{-1} .$$

In terms of inference $\mathrm{var}(\hat{\boldsymbol{\tau}})$ is replaced by

$$\hat{\Omega}(\hat{\boldsymbol{\tau}}) \;=\; \left[\sum_{i=1}^{M} \mathbf{D}_i^T(\hat{\boldsymbol{\beta}}, \hat{\theta}_i)\, \mathbf{V}_i^{-1}(\hat{\boldsymbol{\beta}}, \hat{\theta}_i)\, \mathbf{D}_i(\hat{\boldsymbol{\beta}}, \hat{\theta}_i) \right]^{-1} .$$

To protect against possible misspecification of $\mathbf{V}_i(\boldsymbol{\beta},\theta_i)$ one can use, if necessary, robust inference based on the robust estimator suggested by Liang and Zeger [3],

$$\hat{\Omega}_R(\hat{\boldsymbol{\tau}}) \;=\; \hat{\Omega}(\hat{\boldsymbol{\tau}}) \left[\sum_{i=1}^{M} \mathbf{U}_i(\hat{\boldsymbol{\beta}},\hat{\theta}_i)\,\mathbf{U}_i^T(\hat{\boldsymbol{\beta}},\hat{\theta}_i)\right] \hat{\Omega}(\hat{\boldsymbol{\tau}}) \;,$$

where

$$\mathbf{U}_i(\hat{\boldsymbol{\beta}},\hat{\theta}_i) \;=\; \mathbf{D}_i^T(\hat{\boldsymbol{\beta}},\hat{\theta}_i)\,\mathbf{V}_i^{-1}(\hat{\boldsymbol{\beta}},\hat{\theta}_i)\left(\mathbf{y}_i - \boldsymbol{\mu}_i(\hat{\boldsymbol{\beta}},\hat{\theta}_i)\right) \;.$$

3.3. Computational issues and model linearization

To optimize the objective function (3.2), it is advisable, in practical and computational terms, to find a linearization of the model that transforms the expected value of the variable in a linear function of the parameters $\boldsymbol{\beta}$, as it simplifies the objective function and the variance-covariance matrix considered in it.

Let $\mu_{ij} = E(Y_{ij}) = K(\theta_i)\exp\{\mathbf{x}_j^T\boldsymbol{\beta}\}$ and $\eta_{ij} = \log(\mu_{ij})$. Consider the new random variable

$$\zeta_{ij} \;=\; \eta_{ij} - \log\left[K(\theta_i)\right] + (Y_{ij}-\mu_{ij})\,\frac{d\eta_{ij}}{d\mu_{ij}} \;;$$

then

$$E(\zeta_{ij}) \;=\; \eta_{ij} - \log\left[K(\theta_i)\right] \;=\; \mathbf{x}_j^T\boldsymbol{\beta} \;,$$

which is linear in $\boldsymbol{\beta}$.

Or

$$\begin{aligned}
\zeta_{ij} \;&=\; \mathbf{x}_j^T\boldsymbol{\beta} \,+\, (Y_{ij}-\mu_{ij})\times\frac{1}{\mu_{ij}} \\
&=\; \mathbf{x}_j^T\boldsymbol{\beta} \,+\, \frac{Y_{ij}}{K(\theta_i)\exp\{\mathbf{x}_j^T\boldsymbol{\beta}\}} - 1 \;.
\end{aligned}$$

Let $\boldsymbol{\zeta}$ be the $N\times 1$ vector, $\boldsymbol{\zeta} = \left[\boldsymbol{\zeta}_1^T\;\boldsymbol{\zeta}_2^T\;...\;\boldsymbol{\zeta}_M^T\right]^T$, $\boldsymbol{\zeta}_i = \left[\zeta_{i1}\;\zeta_{i2}\;...\;\zeta_{in_i}\right]^T$, $i=1,...,M$ and $\mathbf{W}$ the block diagonal variance-covariance matrix in $\boldsymbol{\zeta}$, $\mathbf{W} = \bigoplus_{i=1}^{M}\mathbf{W}_i$, where $\mathbf{W}_i$ is a matrix $n_i\times n_i$, symmetric, with generic element w_{jk}. For each group i, $i=1,...,M$ and $j=1,...,n_i$,

$$\begin{aligned}
w_{jj} \;&=\; \mathrm{var}(\zeta_{ij}) \\
&=\; \left[\frac{1}{K(\theta_i)\exp\{\mathbf{x}_j^T\boldsymbol{\beta}\}}\right]^2 \mathrm{var}(Y_{ij}) \\
&=\; \frac{C(\theta_i)}{\left[K(\theta_i)\right]^2} + \frac{1}{K(\theta_i)\exp\{\mathbf{x}_j^T\boldsymbol{\beta}\}} \;.
\end{aligned}$$

On the other hand, for $j \neq k$ in the i group,

$$\begin{aligned} w_{jk} &= \operatorname{cov}(\zeta_{ij}, \zeta_{ik}) \\ &= \frac{\operatorname{cov}(Y_{ij}, Y_{ik})}{\Big[K(\theta_i)\exp\{\mathbf{x}_j^T \boldsymbol{\beta}\}\Big]\Big[K(\theta_i)\exp\{\mathbf{x}_k^T \boldsymbol{\beta}\}\Big]} \\ &= \frac{C(\theta_i)}{[K(\theta_i)]^2} \ . \end{aligned}$$

The minimization problem (3.2) becomes equivalent to,

$$\min(\boldsymbol{\zeta} - \mathbf{X}\boldsymbol{\beta})^T\,\mathbf{W}^{-1}(\boldsymbol{\zeta} - \mathbf{X}\boldsymbol{\beta}) \ , \tag{3.4}$$

where $\mathbf{X}$ is a model matrix of order $N{\times}p$, $\boldsymbol{\zeta}$ is a $N{\times}1$ vector, $\boldsymbol{\zeta} = [\boldsymbol{\zeta}_1^T\ \boldsymbol{\zeta}_2^T\ ...\ \boldsymbol{\zeta}_M^T]^T$, $\boldsymbol{\zeta}_i = [\zeta_{i1}\ \zeta_{i2}\ ...\ \zeta_{in_i}]^T$, $i = 1, ..., M$ and $\mathbf{W} = \bigoplus_{i=1}^{M} \mathbf{W}_i$, $\mathbf{W}_i = [w_{jk}]_{j,k=1,...,n_i}$, with

$$\begin{aligned} w_{jj} &= \frac{C(\theta_i)}{[K(\theta_i)]^2} + \frac{1}{K(\theta_i)\exp\{\mathbf{x}_j^T\boldsymbol{\beta}\}} \ , \\ w_{jk} &= \frac{C(\theta_i)}{[K(\theta_i)]^2} \ , \qquad j \neq k \ . \end{aligned}$$

The following algorithm is proposed.

Algorithm:

1. Let $t = 0$. Obtain initial estimates for $\boldsymbol{\beta}$, $\hat{\boldsymbol{\beta}}^{(0)}$.

   A log-linear model considering all variables as independent can be used, so that,

   $$\log \boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta} \ ,$$

   where $\boldsymbol{\mu} = E(\mathbf{Y})$, $\mathbf{Y}$ is the $N{\times}1$ vector of all variables, each obeying a Poisson law with mean μ_{ij}, $i{=}1, ..., M$, $j{=}1, ..., n_i$, $\mathbf{X}$ is a model matrix of order $N{\times}p$, and $\boldsymbol{\beta}$ is a $p\times 1$ vector of unknown parameters to be estimated, considering in $\boldsymbol{\beta}$ all the main effects of the model. Thereby $\hat{\boldsymbol{\beta}}^{(0)}$ is found and it will be used in 4.

2. Obtain initial estimates for θ_i, $\hat{\theta}_i^{(0)}$, $i{=}1, ..., M$.

   The estimates can be initialized near zero, or can be obtained by finding the Ordinary Least Squares estimates $\hat{\theta}_i$, that minimizes the objective function

   $$\sum_{i=1}^{M}\Big(\mathbf{y}_i - K(\theta_i)\exp\{\mathbf{X}_i\hat{\boldsymbol{\beta}}^{(0)}\}\Big)^T\Big(\mathbf{y}_i - K(\theta_i)\exp\{\mathbf{X}_i\hat{\boldsymbol{\beta}}^{(0)}\}\Big) \ ,$$

   where $\hat{\boldsymbol{\beta}}^{(0)}$ was found in 1.

3. Compute $K_i^{(t)}\!=\!K(\hat{\theta}_i^{(t)})$, $C_i^{(t)}\!=\!C(\hat{\theta}_i^{(t)})$, following (3.1) and also $A_i^{(t)}\!=\!\frac{C_i^{(t)}}{(K_i^{(t)})^2}$, $i\!=\!1,...,M$.

4. Compute

$$\hat{\zeta}_{ij}^{(t)} = \mathbf{x}_j^T\hat{\boldsymbol{\beta}}^{(t)} + \frac{y_{ij}}{K_i^{(t)}\exp\{\mathbf{x}_j^T\hat{\boldsymbol{\beta}}^{(t)}\}} - 1 \ , \qquad i=1,...,M, \quad j=1,...,n_i \ ,$$

$$\hat{\mathbf{W}}_i^{(t)} = \mathbf{J}_{n_i}A_i^{(t)} + \operatorname{diag}\left\{\frac{1}{K_i^{(t)}\exp\{\mathbf{X}_i\hat{\boldsymbol{\beta}}^{(t)}\}}\right\} \ , \qquad i=1,...,M \ ,$$

(where $\mathbf{J}_{n_i}$ is a square n_i dimensional matrix of ones and $\mathbf{X}_i\boldsymbol{\beta}$ is a $n_i\!\times\!1$ vector with elements $\mathbf{x}_j^T\boldsymbol{\beta}$, $j\!=\!1,...,n_i$),

$$\hat{\mathbf{W}}^{(t)} = \operatorname{diag}\left\{\hat{\mathbf{W}}_1^{(t)},...,\hat{\mathbf{W}}_M^{(t)}\right\} \ ,$$

and

$$\hat{\Sigma}^{(t)} = \left[\hat{\mathbf{W}}^{(t)}\right]^{-1} \ .$$

5. Update $\hat{\boldsymbol{\beta}}^{(t+1)}$ and $\hat{\theta}_i^{(t+1)}$ that minimize

$$(\boldsymbol{\zeta}-\mathbf{X}\boldsymbol{\beta})^T\ \hat{\Sigma}^{(t)}(\boldsymbol{\zeta}-\mathbf{X}\boldsymbol{\beta}) \ ,$$

where $\mathbf{X}$ is a model matrix of order $N\times p$, $\boldsymbol{\zeta}$ is a $N\times 1$ vector, $\boldsymbol{\zeta}=\left[\boldsymbol{\zeta}_1^T\ \boldsymbol{\zeta}_2^T\ ...\ \boldsymbol{\zeta}_M^T\right]^T$, $\boldsymbol{\zeta}_i\!=\![\zeta_{i1}\ \zeta_{i2}\ ...\ \zeta_{in_i}]^T$, $i=1,...,M$.

6. Let $t=t+1$. Iterate steps 3 to 6 until the estimates have all stabilized.

Notice that the algorithm uses the IRGLS estimation.

In the final model the fitted values are given by

$$\hat{y}_{ij} = K(\hat{\theta}_i)\exp\{\mathbf{x}_j^T\hat{\boldsymbol{\beta}}\} \ , \qquad i=1,...,M, \quad j=1,...,n_i \ .$$

Note the i-group effect $K(\theta_i)$ present in the fitted values.

In summary, in this proposed modelling strategy, the starting point is a conditional model in $\mathbf{Y}_i|b_i$, considering $\log\left[E(\mathbf{Y}_i|b_i)\right]=\mathbf{X}_i\boldsymbol{\beta}+\mathbf{1}_{n_i}b_i$. A distribution for the random variable b_i is introduced that allows correlation structure representation within the groups. The parameters are then estimated using the IRGLS method, based on $\mathbf{Y}$ moments.

4. A MODELLING EXAMPLE WITH WATER SAMPLES

The total number of coliforms (rod-shaped bacteria) in a water sample is measured in MPN/100ml, number of coliforms (in thousands) per 100 ml of water.

A set of grouped data is analyzed here. The number of coliforms in three collection spouts was registered in Lis river of the Leiria district, Portugal, in 54 occasions [source: INAG, Portugal].

The data is presented in the following graphics by *temperature* and *pH* which are the covariates of the modelling process.

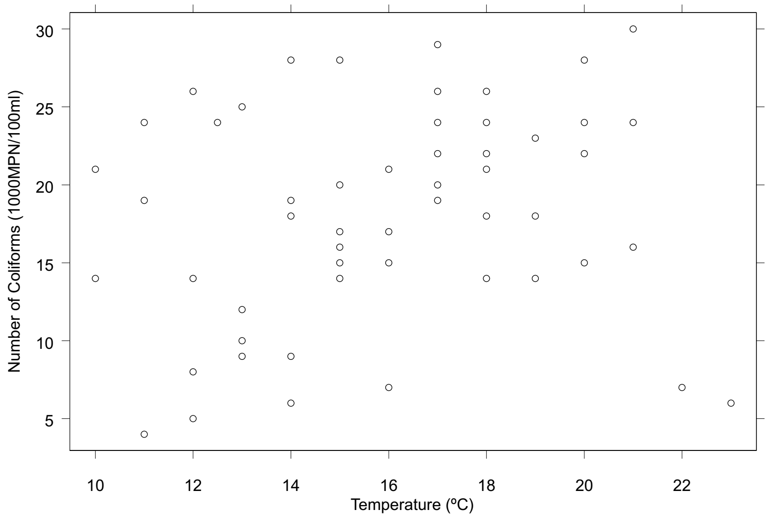

Figure 1: Number of coliforms by temperature.

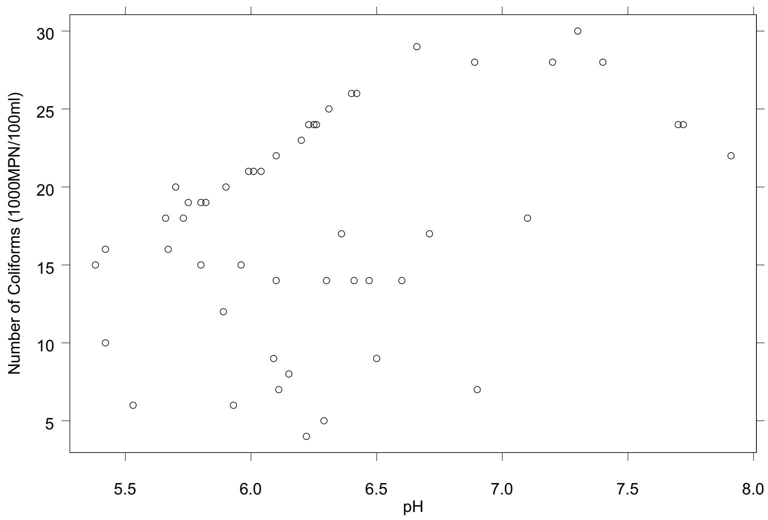

Figure 2: Number of coliforms by pH.

Observing the earlier graphics no systematic pattern is observed. However, looking at Figure 3, which represents the same observations per group — Amor, Milagres and Ponte das Mestras collection spouts, a dependence between the response variable and the covariates is highlighted. It may be also noticed that the response behaves differently for different groups.

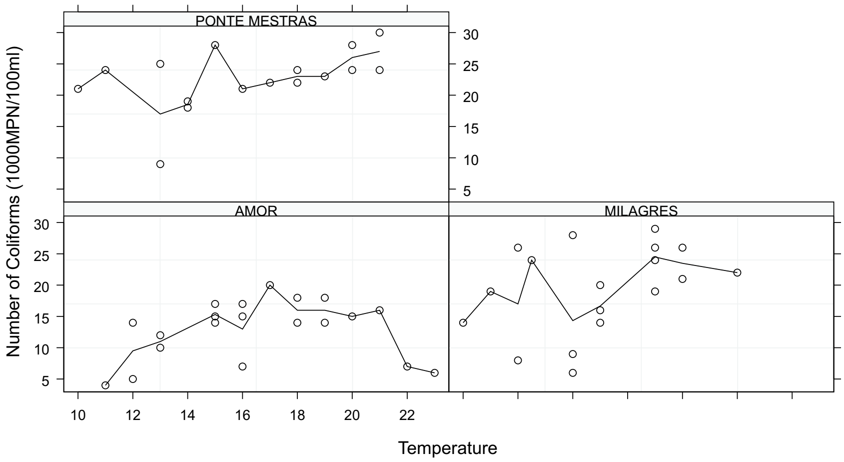

Figure 3: Number of coliforms by temperature and captation.

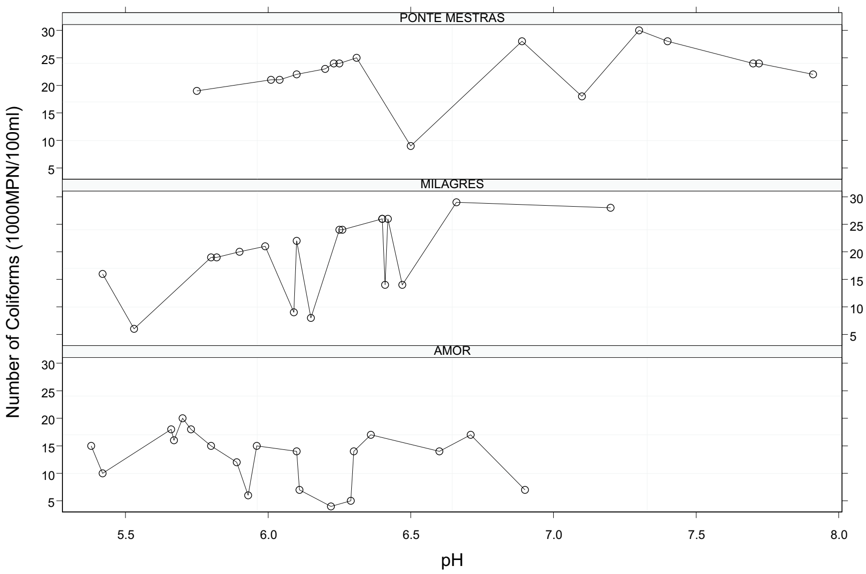

Figure 4: Number of coliforms by pH and captation.

In fact, at Ponte das Mestras and Milagres, the number of coliforms seems to follow the temperature and pH increase. However, at Amor, this is not observed.

The response variable, the number of coliforms, is a discrete variable (counting), suggesting a model based on some Poisson distribution and the method described earlier was implemented. This was done using S language and a small program that supports the method.

The modelling process will start with a log-linear Poisson model (point 1 of the proposed algorithm) considering all the variables as independent, and therefore, $\boldsymbol{\beta}$ initial values are obtained.

Considering the linear predictor

$$\beta_0 + \beta_1\, temp + \beta_2\, pH \ ,$$

$\hat{\beta}_0 = 1.22$, $\hat{\beta}_1 = 0.01$ and $\hat{\beta}_2 = 0.23$ are obtained, where *temp* is the *temperature* covariate. Overdispersion is observed in the model.

The $\theta_i, \ i = 1, 2, 3,$ parameters were initialized near zero.

It was observed that models with an intercept (β_0) have worst convergence, so all the models were considered without this parameter. Starting from $\hat{\beta}_1^{(0)} = 0.02$ and $\hat{\beta}_2^{(0)} = 0.39$, which were obtained from a log-linear Poisson model without intercept, the proposed methodology leads to the estimates

$$\hat{\beta}_1 = 0.03, \quad \hat{\beta}_2 = 0.14, \quad \hat{\theta}_1 = 0.77, \quad \hat{\theta}_2 = 0.98 \quad \text{and} \quad \hat{\theta}_3 = 1.00 \ ,$$

where θ_1 comes from Amor, θ_2 from Milagres and θ_3 from Ponte das Mestras.

However the $\hat{\beta}_1$ and $\hat{\beta}_2$ standard errors were estimated as 0.02 and 0.09, respectively, so they are not jointly significant. The θ_i standard errors were all significant.

So the models whose linear predictor has only one covariate, *temperature* or pH, will be compared.

Model with linear predictor	**Objective function (3.4) value**
$\beta_1\, temp$	78.10
$\beta_2\, pH,$	81.77

The model with the *temperature* covariate is chosen, as it has a lower value for function (3.4). The following estimates and standard errors were obtained in the selected model.

Parameter	**Referred to**	**Estimate**	**Standard Error**
β_1	temperature	0.04	0.01
θ_1	Amor	1.16	0.16
θ_2	Milagres	1.49	0.13
θ_3	Ponte das Mestras	1.48	0.14

The normalized residuals are concentrated in $[-2.04, 1.16]$.

It can be noticed that the water temperature influences the number of coliforms, because the coefficient of the *temperature* covariate is significant, although it has a low estimate ($\hat{\beta}_1 = 0.04$). The number of coliforms increases with water temperature, but not in the same way in all the spouts. In fact, in Amor this is not evident, thereby the correspondent θ_i estimate is the lower one. Probably, in this group, there are some other factors important to the coliform concentrations that were not considered here.

The select quasi-log-linear model, based on the quasi-likelihood function (as overdispersion is present), has linear predictor $\beta_0 + \beta_2\, pH$, considering pH the most significant covariate, but this model has no better fit than the mixed Poisson–Poisson considered in this paper.

As a result, clusters in data should not be ignored. It is possible to model grouped count data with the mixed Poisson–Poisson model and the algorithm proposed above. This methodology estimates the fixed and covariance parameters respecting the between groups correlations structure. Using the IRGLS method it becomes possible to obtain consistent estimates.

REFERENCES

[1] Goldstein, H. (1995). *Multilevel Statistical Models*, 2^a ed., Arnold and Oxford University Press.

[2] Laird, N.M. and Ware, J.H. (1982). Random-effects models for longitudinal data, *Biometrics*, **38**, 963–974.

[3] Liang, K.-Y. and Zeger, S.L. (1986). Longitudinal data analysis using generalized linear models, *Biometrika*, **73**, 13–22.

[4] McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Models*, 2^a ed., Chapman & Hall, London.

[5] McCulloch, C.E. and Searle, S.R. (2001). *Generalized, Linear and Mixed Models*, John Wiley & Sons.

[6] Pinheiro, J.C. and Bates, D.M. (2000). *Mixed-Effects Models in S and S-Plus*, Springer-Verlag.

[7] Vonesh, E.F. and Chinchilli, V.M. (1997). *Linear and Nonlinear Models for the Analysis of Repeated Measurements*, Marcel Dekker.