1 Matrix Formulation

There are matrix (i.e multivariate) formulations of both fixed effects models and random effects models. Brown and Prescott (1999) remarks that the matrix notation makes the underlying theory of mixed effects models much easier to work with. The fixed effects models can be specified as follows;

$$\mathbf{Y} = \mathbf{Xb} + \mathbf{e} \tag{1}$$

$\mathbf{Y}$ is the vector of n observations, with dimension $n \times 1$. $\mathbf{b}$ is a vector of fixed p effects, and has dimension $p \times 1$. It is composed of coefficients, with the first element being the population mean. For the skin tumour example, with the three specified fixed effects, $p = 4$. $\mathbf{X}$ is known as the design 'matrix', model matrix for fixed effects, and comprises 0s or 1s, depending on whether the relevant fixed effects have any effect on the observation is question. $\mathbf{X}$ has dimension $n \times p$. $\mathbf{e}$ is the vector of residuals with dimension $n \times 1$.

The random effects models can be specified similarly. $\mathbf{Z}$ is known as the 'model matrix for random effects', and also comprises 0s or 1s. It has dimension $n \times q$. $\mathbf{u}$ is a vector of random q effects, and has dimension $q \times 1$.

$$\mathbf{Y} = \mathbf{Zu} + \mathbf{e} \tag{2}$$

Again, once the component fixed effects and random effects components are considered, progression to a mixed model formulation is a simple step. Further to Laird and Ware (1982), it is conventional to formulate a mixed effects model in matrix form as follows:

$$\mathbf{Y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{e} \tag{3}$$

($E(\mathbf{u}) = 0$, $E(\mathbf{e}) = 0$ and $E(\mathbf{y}) = \mathbf{Xb}$)

2 Statement of the LME model

A linear mixed effects model is a linear mdoel that combined fixed and random effect terms formulated by Laird and Ware (1982) as follows;

$$Y_i = X_i\beta + Z_ib_i + \epsilon_i$$

- Y_i is the $n \times 1$ response vector

- X_i is the $n \times p$ Model matrix for fixed effects

- β is the $p \times 1$ vector of fixed effects coefficients

- Z_i is the $n \times q$ Model matrix for random effects

- b_i is the $q \times 1$ vector of random effects coefficients, sometimes denoted as u_i

- ϵ is the $n \times 1$ vector of observation errors

The linear mixed effects model is given by

$$Y = X\beta + Zu + \epsilon \tag{4}$$

2.1 Stating the LME Model

The general linear mixed model is

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

where Y is a $(n \times 1)$ vector of observed data, X is an $(n \times p)$ fixed-effects design or regressor matrix of rank k, Z is a $(n \times g)$ random-effects design or regressor matrix, u is a $(g \times 1)$ vector of random effects, and ε is an $(n \times 1)$ vector of model errors (also random effects). The distributional assumptions made by the MIXED procedure are as follows: is normal with mean 0 and variance G; ε is normal with mean 0 and variance R; the random components u and ε are independent. Parameters of this model are the fixed-effects and all unknowns in the variance matrices G and R. The unknown variance elements are referred to as the covariance parameters and collected in the vector *theta*.

The concept of critiquing the model-data agreement applies in mixed models in the same way as in linear fixed-effects models. In fact, because of the more complex model structure, you can argue that model and data diagnostics are even more important. For example, you are not only concerned with capturing the important variables in the model. You are also concerned with distributing them correctly between the fixed and random components of the model. The mixed model structure presents unique and interesting challenges that prompt us to reexamine the traditional ideas of influence and residual analysis. This paper presents the extension of traditional tools and statistical measures for influence and residual analysis to the linear mixed model and demonstrates their implementation in the MIXED procedure (experimental features in SAS 9.1). The remainder of this paper is organized as follows. The Background section briefly discusses some mixed model estimation theory and the challenges to model diagnosis that result from it.

2.2 Laird Ware Formulation

$$\boldsymbol{y_i} = \boldsymbol{X_i}\boldsymbol{\beta} + \boldsymbol{Z_i}\boldsymbol{b_i} + \boldsymbol{\epsilon_i}, \qquad i = 1, \dots, 85$$

$$\boldsymbol{Z_i} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Psi}), \qquad \boldsymbol{\epsilon_i} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\sigma^2 \Lambda})$$

2

2.3 Stating the LME Model

The general linear mixed model is

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

where Y is a $(n \times 1)$ vector of observed data, X is an $(n \times p)$ fixed-effects design or regressor matrix of rank k, Z is a $(n \times g)$ random-effects design or regressor matrix, u is a $(g \times 1)$ vector of random effects, and ε is an $(n \times 1)$ vector of model errors (also random effects). The distributional assumptions made by the MIXED procedure are as follows: is normal with mean 0 and variance G; ε is normal with mean 0 and variance R; the random components u and ε are independent. Parameters of this model are the fixed-effects and all unknowns in the variance matrices G and R. The unknown variance elements are referred to as the covariance parameters and collected in the vector *theta*.

The concept of critiquing the model-data agreement applies in mixed models in the same way as in linear fixed-effects models. In fact, because of the more complex model structure, you can argue that model and data diagnostics are even more important. For example, you are not only concerned with capturing the important variables in the model. You are also concerned with distributing them correctly between the fixed and random components of the model. The mixed model structure presents unique and interesting challenges that prompt us to reexamine the traditional ideas of influence and residual analysis. This paper presents the extension of traditional tools and statistical measures for influence and residual analysis to the linear mixed model and demonstrates their implementation in the MIXED procedure (experimental features in SAS 9.1). The remainder of this paper is organized as follows. The Background section briefly discusses some mixed model estimation theory and the challenges to model diagnosis that result from it.

References

Brown, H. and R. Prescott (1999). *Applied Mixed Models In Medicine*. John Wiley and Sons.

Laird, N. M. and J. H. Ware (1982). Random-effects models for longitudinal data. *Biometrics 38*(4), 963–974.