Some Aspects of Longitudinal Data Analysis

Peter J. Ricci

Thesis Submitted for the Degree of Doctor of Philosophy, Department of Statistics, The University of Adelaide, April 26, 1994
Contents

1 Introduction ........................................ 1
  1.1 Background ........................................ 1
    1.1.1 The Profile and Sum of Profiles Model ............... 1
    1.1.2 Analysis of Non-Gaussian Longitudinal Data ............ 6
  1.2 An Outline of the Thesis .......................... 15

2 REML and the Profile Model ........................... 18
  2.1 Introduction ....................................... 18
  2.2 REML and the General Linear Model .................. 18
  2.3 The Profile Model .................................. 23
  2.4 The Canonical Form ................................ 24
  2.5 Fully Parameterized Covariance Matrix ............... 25
  2.6 The Sum of Profiles Model .......................... 26

3 The Profile Model with Maximum Likelihood and REML estimation ... 30
  3.1 Introduction ....................................... 30
  3.2 Estimation when the covariance matrix is unknown .......... 31
  3.3 Structured Covariance Matrices ...................... 33
    3.3.1 Rao's Covariance Form ......................... 33
    3.3.2 The General Case ............................ 35
  3.4 A Simulation Study: First Order Autoregressive Covariance Structures .......................... 37
  3.5 Discussion ......................................... 41
4 Models for Repeated Categorical Response ........................................... 51
  4.1 Introduction .................................................................................. 51
  4.2 The Linear Model and Inference .................................................... 53
  4.3 Order Selection of Polynomial Models ......................................... 56
  4.4 Oviposition of flies example .......................................................... 58
  4.5 Discussion ..................................................................................... 61

5 Quasi-likelihood and Generalized Estimating Equations ...................... 63
  5.1 Quasi-likelihood ........................................................................... 63
  5.2 Extended Quasi-likelihood ............................................................ 65
    5.2.1 Specifying the Extended Quasi-likelihood ............................... 69
    5.2.2 Indexed Variance Functions .................................................... 66
    5.2.3 A Variable Dispersion Parameter ............................................ 66
  5.3 Marginal Means and First Order Generalized Estimating Equations .... 68
    5.3.1 Extension of the Quasi-likelihood Equations for Longitudinal Data 88
    5.3.2 GEE for the Mean Parameters ................................................. 69
  5.4 A Second Order Extension of the GEE .......................................... 72
  5.5 Estimating Equations by a Case of Quadratic Models ...................... 75
  5.6 Efficiency and Consistency of GEE1 and GEE2 .............................. 77

6 Extensions of the Generalized Estimating Equation Method ................ 79
  6.1 Introduction .................................................................................. 79
  6.2 Occasion-Specific Parameters and Sum of Profiles ......................... 80
  6.3 A Dispersion Model ..................................................................... 84
  6.4 Pseudo-likelihood Estimates ........................................................ 88
  6.5 Estimation of the GEE parameters ................................................. 88
  6.6 Moment Estimation ..................................................................... 89
  6.7 Model Selection Criteria ............................................................... 90
Summary

The analysis of longitudinal data is a very important problem in a wide range of statistical fields. There has been a great deal of work for continuous (Normal and non-Normal) and discrete data, involving parametric and semi-parametric approaches. In this thesis both the Normal and non-Normal cases are considered and the sum of profiles idea of the growth curve model (Normal theory) is extended to non-Normal data.

Chapter 1 provides a literature review of the area and a more detailed description of the thesis.

Chapter 2 introduces the profile model (growth curve model) in a canonical form. The straightforward extension to the sum of profiles model is discussed. We also describe residual maximum likelihood.

In chapter 3 estimation of the parameters for the profile model is considered under maximum likelihood and residual maximum likelihood. Two versions of residual maximum likelihood are derived and used.

Repeated categorical responses are considered in chapter 4. Extensions to the Stanek and Diehl (1988) [134] linear model are given. Connections with the sum of profiles model are discussed.

The theory of quasi-likelihood and generalized estimating equations is introduced in chapter 5. Second order extensions of the generalized estimating equations are also presented. Estimating equations obtained from a class of quadratic models are also considered.

Time dependency of the mean (and dispersion parameters) is one of the topics ex-
anined in chapter 6. The dispersion parameters are also considered as functions of covariates and are estimated using another set of generalized estimating equations.

Application of the GEE methodology of chapters 5 and 6 to three examples is the focus of chapter 7. Positive and negative aspects of the methods are discussed.

In chapter 8 a data set is examined where generalized estimating equations provide a natural approach for analysis. A number of special nested correlation structures are proposed for this data set and tests for these structures are considered.