Correlated Data Analysis: Modeling, Analytics and Applications

PETER X.-K. SONG

Project Set 5

Project 5.1

Marginal GLM Analysis of the Indonesian Children's Health Study Data

This project is supposed to be done individually.

The objective of this project is to acquire experience on the analysis of longitudinal binary observations through the means of marginal GLM, featuring the application of both GEE and QIF approaches and the related SAS software packages. R or SPlus software may be used to make plots.

The data analysis report requires to include the following items:

- Introduction to data
- Preliminary analysis including various plots and cross-sectional regression at each time point and plot the regression coefficients over time to see if these coefficients are constant.
- Modeling
- Estimation/testing significance
- Interpretation
- Residual analysis
- Evaluation of goodness of fit
- Selection of parsimonious model
- Conclusion and discussion

The data has been introduced in Section 1.3.1 and is available for downloaded from webpage

`http://www.stats.uwaterloo.ca/~song/BOOKLDA.html`

Project 5.2

Marginal GLM Analysis of the Sandhopper Orientation Data

This project is supposed to be done with a team.

The focus of this project is on the analysis of longitudinal angular observations through the means of marginal GLM with von Mises distribution, with the utility of GEE or GEE2 approach. The presentation of this project report is the same as that described in **Project 5.1**. Since there is no software package available to fit the von Mises GEE, the related numerical implementation using R software is necessary and gives rise to the biggest technical challenge of the analysis.

The sandhopper orientation data has been introduced in Section 1.3.4 and can be downloaded from the website of the book. This data is analyzed later in Chapter 7 using the conditional modeling approach.

As a result, this project builds an extension of **Problem 2.5**.