STAT 5300: Analysis of Longitudinal Data

Winter Semester 2006

Instructor: J. Flemming
Office: Chase 122
Telephone: 494-1268
E-mail: Joanna.Flemming@Dal.ca

Lectures: Tuesday and Thursday 1:05-2:25
Room: Chase 227
Office Hours: Tuesday and Thursday 2:30-3:30

Text: Analysis of Longitudinal Data
Peter J. Diggle, Patrick Heagerty, Kung-Yee Liang and Scott L. Zeger

Reference Papers: To be given later. ← will look at some current issues

Course Description: This course is concerned with the theory and application of appropriate statistical techniques for analysis of longitudinal data, data that are collected repeatedly over time on a number of subjects. Topics include generalized estimating equations; fixed, random and mixed effects linear models; generalized linear models, diagnostics and model checking; as well as missing data issues.

Method of Evaluation:

Assignments	20%	(4 @ 5% each)
Project	30%	
Midterm Exam	25%	} each 1.5 hrs.
Final Exam	25% (Note: There will be no supplemental exam)	} each 1.5 hrs.

Grading Scheme:

A+	90 and above
A	85 up to but not including 90
A-	80 up to but not including 85
B+	75 up to but not including 80
B	70 up to but not including 75
B-	65 up to but not including 70
C+	60 up to but not including 65
C	55 up to but not including 60
C-	50 up to but not including 55
D	47 up to but not including 50
F	below 47

Assignments: There will be 4 assignments that will provide students the opportunity to use the techniques discussed in class to analyze data sets using appropriate software packages (S+/R, winBUGS). Late assignments will not be accepted.

Project: Each student will be required to submit (individually) a term project. The student will perform and document a complete analysis of a longitudinal data set of interest to them. Further details to be provided in class. Could turn into a paper!

Week	Dates	Topics	Requirements*
1	January 5	Chapter 1	
2	January 10 and 12	Chapters 2 and 3	
3	January 17 and 19	Chapters 5 and 6 Ho: A1	Assignment 1 due
4	January 24 and 26	Chapters 7 and 8 Ho: A2	
5	February 7 and 9	Chapter 9 Chris teaching.	Assignment 2 due
6	February 14 and 16	Chapter 10 Ho: A3	Feb. 17th 9:30-11 am.
7	February 21 and 23	Project Work	
8	February 28 and March 2		
9	March 7 and 9	Chapter 11	Assignment 3 due
10	March 14 and 16	Chapter 12	
11	March 21 and 23	Chapter 13	Assignment 4 due
12	March 28 and 30	Special Topics	
13	April 4 and 6	Special Topics	Project due

*due dates are on Thursday of week indicated

April 18th < 24th.

Exams: Both the midterm and final exam will be held outside of class time on February 17 and April 10, respectively. The exact times will be determined via discussion on the first day of class.

STAT 5300: Longitudinal Data Analysis

Final Exam - 25 Points Total

April 10 2006

1. *(5 points)* Explain the motivation for proposing marginalized models. Discuss the formulation and interpretation of a marginalized random effects model using the dataset you are investigating for your project.

2. *(5 points)* For a longitudinal data set, let Y^* represent the complete set of measurements which would have been obtained were there no missing values. Next let $Y^* = (Y^{(o)}, Y^{(m)})$ where $Y^{(o)}$ represents those measurements actually obtained and $Y^{(m)}$ represents those that are missing, for whatever reason. Finally, let R be a set of indicator random variables, denoting which are in $Y^{(o)}$ and which are in $Y^{(m)}$. Using this notation provide definitions of (a) completely random, (b) random and (c) informative missingness.

3. *(5 points)* Consider fitting the following random effects model to a longitudinal data set using WinBUGS:

```
for ( i in 1 : N ) {
for ( j in 1 : T ) {
    y[i,j] ~ dbin(p[i,j],1)
    logit(p[i,j]) <- alpha0 + alpha1 * x1[i,j] + alpha2 *x2[i,j] + b[i]
}
b[i]~dnorm(0.0,tau)
}
```

   To complete formulation of the model indicate those random variables for which prior distributions must be specified. Provide an appropriate distributional form for each of these prior distributions.

4. *(5 points)* Explain why the cross-validation MCMC procedure recently proposed by Cantoni et al. is desirable when large numbers of potential explanatory variables must be considered. State one other advantage of the approach over a standard backwards stepwise procedure.

5. *(5 points)* What additional information is provided by the robust gee approach that we studied in this course (as compared with the classical version). In the situation where the robust and classical versions agree quite well for a particular data set of interest, what can one conclude?