

SPPH 501
Analysis of Longitudinal & Correlated Data
(January – April 2024)

TIME & PLACE: Term 2, Tuesdays, 2:00- 5:00 P.M.

LOCATION: SPPH, Room 143

INSTRUCTOR: Dr. Ying C MacNab

OFFICE: Room 414

TELEPHONE: (604) 822-5593

EMAIL: ying.macnab@ubc.ca

OFFICE HOURS:

2-5pm Thursdays

TEACHING ASSISTANT:

TBD

COMPUTER LABS:

Students in 501 receive priority to use the School computing lab on Tuesdays, Noon. - 2P.M. Students are welcome to use their own rather than the lab computers.

COURSE PHILOSOPHY AND OBJECTIVES:

This course will introduce students to concepts and methods in the analysis of correlated data, with special emphasis on longitudinal and hierarchical data. By the end of the course students are expected to:

1. Recognize the types of study/sampling designs that give rise to correlated data.
2. Translate conceptual models relating health outcomes and their determinants into statistical models.
3. Have basic knowledge of linear model with correlated errors, GEE models, generalized linear mixed models, Bayesian hierarchical models, and related methods of estimation and inference
4. Identify different analytic approaches to longitudinal/correlated data analysis and explain the advantages and disadvantages of each approach.

PREREQUISITES:

SPPH 400 and 500, 5002 or their equivalents.

COURSE REFERENCES:

The primary reference used in this course will be

1. Gelman A, and Hill J. Data Analysis Using Regression and Multilevel/Hierarchical Models. 2007. Cambridge.

though some material will also be taken from:

1. Fitzmaurice GM, Laird NM, and Ware JH. 2011. Applied Longitudinal Analysis (2nd Edition). Wiley.
2. Diggle PJ, Heagerty P, Liang KY, and Zeger SL. 2002. Analysis of Longitudinal Data (2nd Edition). Oxford University Press.

Computing:

1. The Gelman text provides some examples in R. Additional materials will be distributed during the course.
2. WinBUGs manual (<http://www.mrc-bsu.cam.ac.uk/bugs>)
3. Jim Albert 2007. Bayesian computation with R

Additional:

1. Goldstein H. 2003. Multilevel Statistical Models (3rd Edition). Arnold. (A pdf of an older version, which suffices for this course, is available for download from http://www.ats.ucla.edu/stat/examples/msm_goldstein/default.htm).
2. Pinheiro JC, and Bates DM. 2000. Mixed Effects Models in S and S-PLUS. Springer. (As indicated by the title, this text uses S-PLUS as the computing environment. While many of the basic functions and commands in R and S-PLUS are identical, this is not the case for mixed effects models so you should not purchase this text for learning the computing. The book is listed here because although it is a text on computing, the examples used also illustrate many of the important concepts and theory important underlying mixed effects models.)

COURSE NOTES:

Course slides will be available on selected topics. Note that these are summary slides, not detailed content. Students are expected to have read the slides and related material prior to the class. In-class time will usually alternate between short presentations of basic ideas by the instructor and class discussion of these ideas.

STATISTICAL COMPUTING:

Statistical analysis in this course will be illustrated using R and WinBUGs. R is a free software environment for statistical computing and can be downloaded through the R Project homepage: www.r-project.org. WinBUGs is a free statistical software for Bayesian analysis using Markov chain Monte Carlo (simulation) methods and can be downloaded at <http://www.mrc-bsu.cam.ac.uk/bugs>.

Students are welcome to use other software packages (SAS, Stata, etc) instead of R to complete the assignments (limited support from the instructor and TA).

COURSE EVALUATION:

The course is graded on a Pass/Fail basis (68% required for a Pass). The components that go into the final grade are:

Assignments: 30%

Final Project: 50% (30% report, 20% class presentation)

Class Participation: 20%

ASSIGNMENTS:

Assignments will be distributed periodically.

The main aim of the assignments is to help students become comfortable with translating conceptual models into mathematical models.

Late assignments are not accepted.

Assignments should be typed or neatly written.

Mathematical notation and language should be used very carefully. Marks will be deducted for improper use of notation/jargon, particularly in cases where it renders the answer vague.

Students working with the software are encouraged to consult with each other about how to get things done but all written work must be completed individually.

Present only those parts of the computer output that are relevant to your answer and highlight or underline the specific items of interest. Alternatively, transcribe those items to another page if you prefer.

FINAL PROJECT:

There is considerable flexibility in the scope of the final project and students are encouraged to propose and develop a project in consultation with the instructor. Examples of projects include:

1. Critiquing of longitudinal analyses used in published studies: Identify the type of design and the models used. How appropriate were the models to the questions of interest? What were the limitations of the design and analysis? What might have been done or could be done in the future to better address the questions?
2. Design of a longitudinal study: Identify a question which requires longitudinal data to answer. Discuss possible ways (and the advantages/disadvantages of each) of collecting data to answer the question. Propose specific models for analyzing the data with discussion of how each model is related to and accounts for various features of the data.
3. Analysis of a dataset: Starting with a given dataset and various questions of interest, perform the necessary analyses (graphical, modeling, diagnostics, interpretation) to address those questions.
4. Special topics: Develop a "mini-lecture" that enhances material on a topic (e.g. missing data, transition models, causal models, etc) that is not discussed in-depth in class.
5. Other software packages: Develop a "tutorial" to show how the analyses illustrated in class using R would be done using another software package. Compare and contrast the capabilities/limitations of the two packages.

EXAMS:

There are no exams for this course.

COURSE OUTLINE:

The following schedule is tentative. Changes may be made to accommodate the needs and interests of the students.

Class #1

Introduction and preliminaries.

Online search for (preferably most recent) public health research papers that involve analysis of correlated data (longitudinal, multilevel, spatial, etc).

Reproducible research

Assignment 1 will be announced

Class #2

Review of distributions, probability, conditional probability, expected value, correlation. Review of sampling designs and data structures. Sources of correlated data (hierarchical, longitudinal, spatial). Implications for statistical inference of various types (association, causality, prediction). Principles of regression modeling. Examples.

Readings: Gelman, Ch 1 – 4, 9

Class #3, Assignment #1 due.

Present of assignment 1

Grouped data. Stratification and clustering. Levels of analysis. Fixed and random effects. Analysis of studies with observations at two time-points. Exploring and summarizing grouped data.

Readings: Gelman, Ch 11. Fitzmaurice, Ch 1 & 2. Diggle, Ch 1.

Class #4: Assignment #2a,b will be announced

Matrices. Correlation structures. Design and sample size considerations. Analytic approaches (marginal models, mixed effects models, Bayesian hierarchical models.

Readings: Gelman, Ch 20. Fitzmaurice, Sec 7.1 - 7.7, Ch 3. Diggle, Ch 2 & 3.

Class #5

Marginal models: linear models with correlated error terms, GEE model, related methods of estimation and inference

Reading materials to be distributed

Class #6:

Linear mixed effects models. Modeling of mean and variance structures. ML and REML estimation. Matrix representation.

Readings: Gelman, Ch 12 & 13. Fitzmaurice, Ch 8. Diggle, Ch 4 & 5.

Class #7: Assignment #2b due.

Midterm (presentation of the term project ideas and outline)

Class #8: Outline of proposed term project (Assignment #2a) due.

Review of generalized linear models (GLMs). Introduction to generalized linear mixed effects models (GLMMs). WLS, GEE, MLE-EM, PQL and MCMC estimation.

Readings: Gelman, Ch 5, 6, 14 & 15. Fitzmaurice, Ch 10 & 12. Diggle, Ch 9. WinBUGs Examples Vol I

Class #9: Assignment #3 will be announced

Multilevel (linear) models: mixed effects model expression, Bayesian hierarchical model expression, MCMC using WinBUGs and R.

Class #10: Draft of projects background/review and data analysis outline due

GLMM and Bayesian GLMM for multilevel data

Readings: WinBUGs Examples Vol I, reading materials to be distributed

Class #11:

Bayesian disease mapping

Readings: GeoBUGs manual and examples, reading materials to be distributed

Class #12:

Assignment #3 due.

Additional topics (continued) as time allows. Student presentations.

Class #13:

Student presentations.