Concept and application of multilevel statistical methods to health, social science and geographical research

Location: *CIQSS, 3535 Queen-Mary, Suite 420, Montréal*
Dates: July 20-23, 2015

Instructors

This training session will be under the responsibility of:

Dr. S V Subramanian
Professor of Population Health and Geography
Department of Social and Behavioral Sciences
Harvard T.H. Chan School of Public Health
Email: svsubram@hsph.harvard.edu

Dr. Alexandre Lebel
Professor at the Graduate School of Land Management and Regional Planning, Laval University
Evaluation Platform on Obesity Prevention
Quebec Heart and Lung Institute
Email: alexandre.lebel@esad.ulaval.ca

Participants

The course is open to graduate students and postdoctoral fellows as well as to professors and practicing researchers. Participants should have some knowledge of multilevel modeling OR a reasonable level of proficiency in concepts related to linear regression and variance.

Course Objectives

By the end of the course, you will be able to:

- Recognize a research problem requiring a multilevel modeling approach.
- Describe the technical and substantive advantages of multilevel models.
- Manage data for statistical modeling of hierarchic structures.
- Explain the basic principles of multilevel modeling using graphical, verbal and statistical language.
- Implement modern methods for understanding clustering and heterogeneity.
- Develop a variety of models that enable quantitative assessment of contextual effects.
- Interpret and communicate the meaning of results for simple and complex linear multilevel models.
- Utilize *MLwiN* procedures to perform multilevel analyses.
- Evaluate research papers that apply multilevel models.
- Apply multilevel models to a research problem according to a well-articulated research strategy.

Course Description

This course is designed to provide you with training in the concept and application of multilevel statistical modeling. You will be motivated to think about correlated and dependent data structures that arise due to sampling design and/or are inherent in the population (such as pupils nested within schools; patients nested within clinics; individuals nested within neighborhoods and so on).

The substantive purpose of this course is to enable quantitative assessments of the role of contexts (*e.g.*, schools, clinics, neighborhoods) in predicting individual outcomes. This will be accomplished by developing a range of multilevel models along with a detailed discussion of the statistical properties and the interpretation of each model. The course would give a thorough consideration of normal theory two-level linear models along with an introduction to more advanced topics including three-level structures, multilevel logistic models, estimation properties as well as multiple membership and cross-classified hierarchies. Empirical presentations will focus on multilevel analysis using *MLwiN* – a specialized software to handle models with complex data structures.

Materials

All material will be provided in class. This includes all presentation slides and a text book: *Jones K., Subramanian S V,* Multilevel statistical models: concepts and applications*. Boston, MA: Harvard T.H. Chan School of Public Health, 2015.*

Training plan

The course will combine lectures on the fundamentals of multilevel modeling where participants would get a chance to specify, interpret and display multilevel linear regression models. The general pattern for the lectures would be to introduce the different methodological concepts, graphically, followed by their statistical specification. This would be followed by estimating multilevel models using 'real data sets' and interpreting them.

In class schedule

The course is scheduled from **9:15 am to 4:00 pm.** All classes and material will be given in English although assistance is also available in French.

8:45 to 9:15 AM- access to laboratory for individual work
9:15 to 12:00 AM- theory session in class

-Lunch time-

1:00 to 3:00 PM- theory in class
3:00 to 4:00 PM- work in laboratory with technical assistance
4:00 to 6:00 PM- access to laboratory for individual work

Content

Hierarchies and varying relationships

This session will motivate and illustrate the fact that most social science and public health research problems are intrinsically *multilevel* and will consider modeling data simultaneously at different levels. The conceptual and methodological need to recognize and maintain the different levels in quantitative modeling will be discussed. The concept of hierarchies and levels will be implemented using a range of graphical models. A two-level hierarchical model will be introduced to appreciate the idea of varying relationships, with an emphasis on the notion of contextuality. The program to implement multilevel ideas, *MLwiN*, will be introduced.

Random-intercepts and slopes model

This session will consider the statistical implementation of the graphical models outlined in the previous session. Three types of multilevel models – a "null two-level random-intercepts" model, a "random intercepts" model and a "random slopes" model - will be introduced and developed. The characteristics of these models will be discussed. Participants will be familiarized with models that disentangle the two sources of variation – "compositional" and "contextual". The concept of higher-level residuals will be introduced and their utility in comparing contextual differences will be discussed.

Comparing fixed and random-effects models

This session will empirically and algebraically compare the fixed effects and random effects approach to modeling higher levels.

Modeling heterogeneity and variance

This session will demonstrate that multilevel models are fundamentally concerned with estimating variance functions at each level. The different specifications (such as constant, linear and quadratic) that could be used to characterize variance functions will be introduced and discussed. The procedures related to calculating simple and complex auto-correlation function (or Variance Partitioning Coefficient) will be outlined. Besides discussing the empirical applicability of the different variance functions, the relevance of such procedures for exploration and indeed model diagnosis will be discussed. The discussion on modeling heterogeneity as a variance function will be restricted in this session to continuous predictor/independent variables.

Readings: Chapter 12, 15

Modeling categorical predictors

The multilevel principles and procedures discuss thus far will be applied in this session to models with categorical predictors. A range of models including dummy variables and interactions in the fixed part; level-1 categorical predictors allowed to vary at level-2 will be specified and their interpretation discussed.

Readings: Chapter 17-18

Modeling higher-level predictors

This session will introduce models with predictor variables measured at the higher-level. The complex association between individual outcomes and contextual predictors will be implemented using multilevel procedures. The technical and substantive advantages of multilevel techniques in handling contextual predictors will be outlined. In terms of implementation, the session will show how to replicate data from an input file containing higher level variables, how to construct higher level variables based on individual measures and how to estimate and display 'cross-level interactions'.

Readings: Chapter 19-20