UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

SYLLABUS DEL CORSO

StatisticAlps 2019-2020

1920-StatisticAlp

Aims

THE ANALYSIS OF LONGITUDINAL AND LIFE HISTORY DATA

Richard Cook - Professor of Statistics in the Department of Statistics and Actuarial Science at the University of Waterloo, Canada

Daniel Farewell - Reader of Statistics in the School of Medicine at the Cardiff University in UK

Statistic Alps is a residential training course on advanced statistical issues of interest in clinical research. It is traditionally held in Ponte di Legno, a place in the Alps with beautiful surroundings. The course has reached the 9th edition and this year will be in a winter format.

The general aim of 2020 edition is to provide an introduction to statistical methods for the analysis of longitudinal and life history data. An emphasis will be given to the kinds of data arising in epidemiology and public health research, with some issues being specific to the analysis of data from clinical studies. Substantive examples from medical science will be used throughout the course to motivate the methods and illustrate the different interpretations given to estimates of intervention and other covariate effects. R code and selected output will be provided in worked examples.

Contents

We will begin with a focus on common approaches for the analysis of repeated measurements from individuals over common scheduled assessment times, including mixed effects models, generalized estimating equations, and autoregressive models. Models and methods will then be discussed for the analysis of life history data obtained from continuous observation of individuals who are subject to right-censoring. The assumptions justifying the various approaches to analysis will be highlighted, and the interpretation of covariate effects and other possible estimands will be emphasized. Recurring themes will include robustness, the implications of a dependence between the longitudinal or life history process and the observation process (i.e. missing data, censoring and informative observation mechanisms), and causal inference. Substantive examples from medical science will be used throughout the course to motivate the methods and illustrate the different interpretations given to estimates of intervention and other covariate effects. R code and selected output will be provided in worked examples.

Detailed program

1ST DAY – 2 MARCH 2020
LONGITUDINAL DATA ANALYSIS VIA HIERARCHICAL AND MARGINAL MODELS

- 9.00:13.00
The assumptions underlying the formulation of hierarchical random effects models will be covered along with the construction of likelihoods;
software will be discussed and applied to illustrative examples.

- 14.30:16.00 and 16.30:18.30
Robustness properties of marginal models and the theory of generalized estimating equations will then be covered with applications given to illustrate how to fit such semiparametric methods; the differences in the interpretation of estimators from the random effect and marginal formulations will be contrasted.

2ND DAY – 3 MARCH 2020
TRANSITIONAL ANALYSIS OF LONGITUDINAL DATA

- 9.00:13.00
The formulation of models and likelihoods will be discussed based on conditional (autoregressive) analyses of continuous and discrete data in discrete time; the latter will introduce Markov and semi-Markov models and generalizations

- 14.30:16.00 and 16.30:18.30
Continuous time processes will be considered for categorical data (i.e. discrete state spaces). Intensity functions will be defined and likelihoods will be constructed in a general way for right censored data.

3RD DAY – 4 MARCH 2020
RECURRENT EVENT AND MULTISTATE MODELS

- 8.30:12.30
Particular types of multistate processes will be covered including recurrent event, competing risk, illness-death, and more general multistate processes. Software will be presented and the interpretation, strengths, and limitations of analyses based on intensity functions, rate functions and other marginal features; these will be discussed from both the conceptual standpoint and through applications.

4TH DAY – 5 MARCH 2020
CHALLENGES WITH INCOMPLETE DATA

-9.00:12.30
The general theme for this day is the challenge of incomplete data arising from intermittent observation of continuous-time processes. Methods for the analysis of longitudinal data will be covered in which the observation times are irregularly spaced.

- 14.00:16.00
Methods for fitting multistate models under intermittent observation will then be discussed with a particular emphasis on Markov models. The assumptions justifying the use of standard methods will be highlighted along with methods which can accommodate some form of dependence between the response and observation processes.

Prerequisites

Basic knowledge of survival analysis. Students are espected to bring their own laptop with the most recent version of R or R studio installed.

Teaching form

Traditional classes, tutorials using the R sofware, discussion.

Textbook and teaching resource

Slides and code fragments provided by the instructors.

Semester

2 - 5 March 2020

Assessment method

Self assessment test.

Office hours