BIOSTAT 696/896: Spatial Statistics
Winter 2013

Instructor: Veronica J. Berrocal
Office: SPH II, M4525
Email: berrocal@umich.edu

Administrative information
Lectures: Mon-Wed 10-11:30 AM, Room SPHII M1170
No class: Mon 01/21/2013; Mon 03/04/2013; Wed 03/06/2013; Mon 03/11/2013 or Wed 03/13/2013
Office hours: Mon: 12:00-1:00 PM; Tue: 10:00-11:00 AM and by appointment
C-Tool course webpage: BIOSTAT 696/896 W13

Course description
With the advent of new technologies and the availability of geographical information systems (GIS), it has become increasingly common for researchers to collect data that are spatially referenced. As a natural consequence, in recent years there has been a great interest in developing statistical methods to analyze spatial data for a wide variety of applications, from environmental disciplines to public health. This course will introduce the theory and methods of spatial statistics and will discuss methods for inference on spatial processes within a geostatistical and a hierarchical Bayesian framework.

Learning objectives
This course is designed to introduce students in a graduate program to statistical methods for the analysis of spatial data. The course will cover theory and methods developed for the three major branches of spatial statistics: (i) point-referenced data, (ii) areal/lattice data, and (iii) point processes. Students will also be exposed to the computational aspects of spatial statistics, learning how to use R for the analysis of spatial data, and will be exposed to applications and real datasets from environmental sciences and public health.
Competencies covered in this course

Core Competencies:

- Describe preferred methodological alternatives to commonly used statistical models when assumptions are not met. (Partially satisfied)
- Apply common statistical methods for inference. (Fully satisfied)
- Apply descriptive and inferential methodologies according to the type of study design for answering a particular research question. (Partially satisfied)

Biostatistics:

- Develop knowledge to communicate and collaborate effectively with scientists in a variety of health-related disciplines to which biostatistics is applied (e.g. public health, medicine, genetics, biology, psychology, economics, management and policy). (Partially satisfied)
- Become well-versed in the application of core statistical techniques (biostatistical inference, linear regression, generalized linear models, analysis of variance (ANOVA), linear mixed model) and 4-5 selected statistical specialization techniques. (Partially satisfied)
- Select appropriate techniques and apply them to the processing of data from health studies. (Partially satisfied)
- Interpret the results of statistical analysis and convert them into a language understandable to the broad statistical community. (Partially satisfied)
- Develop written and oral presentation skills and other scientific reporting skills, based on statistical analyses for public health, medical and basic scientists and educate lay audiences. (Partially satisfied)
Prerequisites
BIOSTAT 601, 602, 650, 653 or similar master level mathematical statistics and regression courses. Familiarity with the R computing environment is assumed.

Textbook
S. Banerjee, B. P. Carlin and A. E. Gelfand, “Hierarchical modeling and analysis for spatial data”, Chapman & Hall/CRC.

Suggested textbooks
The following textbooks are also excellent books in spatial statistics:


• N. Cressie, “Statistics for spatial data” (Revised Edition), Wiley.


Topics
The course will tentatively cover the following topics.

• Introduction: different types of spatial data, spatial dependence

• Point-referenced data
  – Exploratory data analysis
  – Models for spatial dependence
  – Spatial interpolation
  – Asymptotics for spatial processes
  – Non-stationarity
  – Bayesian hierarchical models for spatial data
  – Spatial models for non-Gaussian data
• Areal/lattice data
  – Spatial autocorrelation
  – Spatial smoothing
  – Markov random fields
  – Conditionally AutoRegressive (CAR) and Simultaneously AutoRegressive (SAR) models
  – Disease mapping

• Point processes
  – Poisson process
  – Cox process
  – Poisson cluster process
  – Nearest neighbors
  – K-function

Grading
Homework and a final project contribute to the final grade in the following way:

• Homework: 50%
• Final project: 50%

Homework
There will be 4 homework assigned during the class. Homework is due on the day indicated, at the beginning of the class. Late homework will not be accepted except under very special circumstances (no more than one special circumstance per person). Students can discuss about the homework assignment, but the write-up should be done independently. Changing font on someone else’s computer code is not considered “your own work”.
Homework will include both exercises concerned with the derivation of formulas and computational-based problems, such as for example, analysis of a dataset.

Students in the 896 section of this class will be assigned 2/3 additional homework questions on every homework, mostly concerned with more theoretical aspects of spatial statistics. Additionally, students in the 896 section will give a 15-20 minutes presentations based on a reading assigned to the class.

**Final project**

There will be one final project assigned on April 1, 2013 and due at **10 am on April 15, 2013**. Students will present their final project on **April 15-17-22, 2013** during the regular class time and on **April 23, 2013** during the scheduled Final exam time (between 1:30 PM and 3:30 PM).

For students in the 696 section, the final project will involve the analysis of a spatial dataset, chosen among a collection of spatial datasets that will be made available on the C-tools class website. Students should prepare a written report no longer than 7 pages including tables and figures, introducing the motivation of the study, the model used, specifying the method used for inference and presenting the results and conclusions of their analysis.

For students in the 896 section, the final project will involve the analysis of a spatial dataset that the students need to find or provide on their own. The written report should include an introduction, a description of the dataset, a method section, a results and a conclusion/discussion section, with a page limit of 10 pages including tables and figures.

Final projects (both report and presentation) for both students in the 696 and 896 section will be evaluated on clarity of exposition, appropriateness of the model and of the method used for inference.

**Additional requirement for BIOSTAT 896**

The 896 section of this course requires completing a set of additional questions on the homework sets, a more thorough analysis for the final project and a team presentation of an assigned reading.

**Academic integrity**

Students enrolled or registered in a course offered in the School of Public Health are expected to have a conduct consistent with that of a public
health professional. Plagiarism, cheating, use of assistance from other students during written examinations, as well as use of aids of any kind during a written examination, if not approved by the instructor, is considered academic misconduct and will be referred to the Office of Academic Affairs.