Intermediate Spatial Data Analysis
GEOG 5103
Spring 2020

Derby Hall 1080, Monday/Wednesday 11:10-12:30

Instructor
Elisabeth Root, Associate Professor, Geography & Epidemiology
Email: root.145@osu.edu
Office: 1160 Derby Hall
Office Hours: XXX, or by appointment

Teaching Assistant
Sohyun Park
Email: park.2627@osu.edu
Office: 0160 Derby Hall
Office Hours: WR 13:00-15:00, or by appointment

Course Description
OSU Catalog Description: Application of quantitative methods to geographic problems; spatial statistics, area sampling, maps of residuals, regionalization methods, and simulation maps.

My Description: Geography is a diverse discipline with a wide variety of subject matter including physical (environmental), human (socio-economic), and integrated (human-physical) topics of inquiry. Even within physical geography, biogeographers study different phenomena than hydrologists and climatologists. Therefore, it is not surprising to learn that there are a variety of advanced analytical methods that geographers can employ in their studies. Space, and spatial data, complicate traditional statistics and geographic scientists have developed their own statistical tools to properly draw inferences from spatial data. This course focuses regression-based multivariate methods widely used by geographers and other scientists, including linear regression (including generalized linear models) and spatial regression (including spatial autoregressive models, conditional autoregressive models, and geographically weighted regression). By the end of the course, I want you to know how to select the appropriate regression method to answer a research question, be comfortable using multiple software packages to analyze data, and correctly interpret and write-up the results of your statistical analysis.

I don't believe in memorizing formulas or asking students to regurgitate those formulas. Therefore, this course emphasizes hands-on experience and practical/conceptual understanding. You should leave this course with confidence in the methods we have discussed and an appreciation for how these statistical methods are applied to issues in geographic research. The material covered in this course falls into one of four categories: (1) principles of multiple regression, (2) functional forms of the generalized linear model, (3) regression diagnostics, and (4) spatial regression alternatives. Throughout the course, I will be emphasizing the assumptions inherent in regression analyses, consequences of violating these assumptions, and (spatial) solutions when assumptions have been violated.

Course Objectives:

1. To develop "statistical literacy," a working understanding of statistics that can help in critically evaluating data-driven results in the discipline of geography (or urban planning, public health, etc…).
2. To obtain a rich set of statistical tools for data analysis, with an understanding of the how to choose which tool to use and how to implement them in statistical software.
3. To enable you to confidently and carefully interpret the results of data analyses and clearly communicate those results.
4. To receive practical experience in using real datasets to address meaningful research questions.

Course Website: The course schedule, announcements, lecture notes, assignments, readings, datasets, and other course information will be posted on Carmen (https://carmen.osu.edu).

Prerequisites: Students enrolled in this course must have completed an introductory statistics course (e.g. GEOG 4103 or the old GEOG 5100). Introductory courses from other departments are sufficient to meet this requirement.

Text: Chi, G. and J. Zhu (2020). *Spatial Regression Models for Social Sciences.* Los Angeles: Sage Publications.

We will use chapters from the following texts, which will be available on the Carmen site:

Hair, Anderson, Tatham, and Black. (2014). *Multivariate Data Analysis*, 7th Edition. England: Pearson Education.

Gelman, A. and J. Hill. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. New York: Cambridge University Press. (e-book available through the OSU Libraries: https://ebookcentral.proquest.com/lib/ohiostate-ebooks/detail.action?docID=288457)

Additional Recommended Texts:

Bivand, R., E.J. Pebesma and V. Gomez-Rubio (2008). Applied Spatial Data Analysis with R. New York: Springer. (e-book available through the OSU Libraries – you must be on the VPN: http://ebooks.ohiolink.edu/xtf-ebc/view?docId=tei/sv2/9781461476184/9781461476184.xml&query=&brand=default)

Dalgaard, P. (2008). Introductory Statistics with R. 2nd edition. New York: Springer. (e-book available through the OSU Libraries: http://ebooks.ohiolink.edu/xtf-ebc/view?docId=tei/sv/9780387790541/9780387790541.xml&query=&brand=default)

Waller, L. and C. Gotway (2004). Applied Spatial Statistics for Public Health Data. New York: John Wiley. (e-book available through the OSU Libraries: https://ebookcentral.proquest.com/lib/ohiostate-ebooks/detail.action?docID=214360_

Fortin, M-J. and M. Dale (2005). Spatial Analysis: A Guide for Ecologists. Cambridge: University Press.

Software: This course will use the R statistical software environment, which is a free and open source program for statistical computing and graphics (https://www.r-project.org/). There is both a MacOS and Windows version. We will also use R Studio (https://rstudio.com/products/rstudio/), which is a shell for enhanced visualization and programming. These software packages are available on the lab computers, but I recommend you download them and use your own computer. We will be doing some "hands on" work in the classroom, so please bring your laptop to class each day.

Grading: Grades will be based on the following elements:

Homework	4 x 15% =	60% (12.5% each for grads= 50%)
Exams (take home)	2 x 15% =	30%
In-class problem sets (e.g., participation)		10%
Graduate students only:		
Article critique		10%

Late assignments up to 1 week late will be downgraded 20%, 100% thereafter. Students must complete all lab assignments to receive a passing grade, even if they are submitted too late to receive any points.

Homework: There will be four homework assignments, each of which is worth 15% of your overall course grade (12.5% for graduate students). Assignments will be posted by the Wednesday of the week noted in the schedule and due by 5pm the following Tuesday. Each assignment requires that you use R to analyze data, interpret the results of these statistical analyses, and demonstrate an understanding of statistical principles discussed in class. I will not be handing you a “script” for how to do a statistical analysis during the homework. I will provide you with a dataset and some programming advice and ask you to figure out how to use the software packages we learn to run models and answer a set of broad questions. Homework assignments will be submitted electronically via Carmen.

I believe that teamwork is a crucial skill for today's workforce. Therefore, I encourage you to work together on homework assignments. There is a fine line between cooperative work and copying from one another. Please keep in mind that the purpose of this class is for you to understand how to use statistical techniques to analyze data. This goal will be facilitated by working in small groups – not by copying each other’s answers. Therefore, talk and consult with other students as much as you like, but in the end each student is required to complete their own individual written work. If you have any questions or concerns about this distinction, please discuss them with me or the TA prior to turning in your assignment.

Exams: There will be two take-home exams that account for 30% of your overall course grade (midterm = 15%, final = 15%). I will distribute the exams on a Wednesday during class and it must be turned in by Friday at 5pm. Late exams will not be accepted. The content of the examination will include the range of topics covered during the course. In contrast to the homework assignments, the final exam is exclusively a test of individual work; therefore, you are not permitted to work together. Any question regarding the content or format of the exam should be directed to me, not the TA.

All lab/exams must be typed, double-spaced, and use 12-point font. Formulas should be created using an equation editor. Tables should be constructed in Excel; graphs should be generated via R unless otherwise specified by the TA, in which case you will use Excel to generate them. An R file with the code you created for your assignments/exam should also be prepared and submitted with your lab write-up. This requires that you properly annotate your code and save it as a .R file. All assignments will be submitted via Carmen.

Graduate Student Assignment: I will ask graduate students in the class to provide a critical assessment of a journal article of my choosing about 2/3 of the way through the semester. This will county for 10% of your grade. Assessment of other researcher's work is a critically important skill for graduate students, and I will assess whether you can read an article applying statistical principles discussing in class and point out strengths and weaknesses of the researcher's assumptions, implementation of statistical methods, and interpretation of results.

Attendance: Attendance is required to gain an adequate understanding of the course material, complete homework assignments, and pass the exams. You do not need to email me to tell me that you will not be or were not in class on a specific day. However, if you need to miss more than one class session during the course of the semester, you should alert me to the necessity of your absences. It is your responsibility to find out what we covered in class when you were not in attendance. Do this by asking other students in the course or consulting with the TA. Do not email me to ask what we covered in lecture when you were not there.

Additional Policies

Religious Holidays: Please contact me regarding any conflict between religious observance dates and course examinations or assignments.

Disability Statement: The University strives to make all learning experiences as accessible as possible. If you anticipate or experience academic barriers based on your disability (including mental health, chronic or temporary medical conditions), please let me know immediately so that we can privately discuss options. To establish reasonable accommodations, I may request that you register with Student Life Disability Services. After registration, make arrangements with me as soon as possible to discuss your accommodations so that they may be implemented in a timely fashion. SLDS contact information: slds@osu.edu; 614-292-3307; 098 Baker Hall, 113 W. 12th Avenue.

Academic Misconduct: Academic integrity is essential to maintaining an environment that fosters excellence in teaching, research, and other educational and scholarly activities. Thus, The Ohio State University and the Committee on Academic Misconduct (COAM) expect that all students have read and understand the University's Code of Student Conduct, and that all students will complete all academic and scholarly assignments with fairness and honesty. Students must recognize that failure to follow the rules and guidelines established in the University's Code of Student Conduct and this syllabus may constitute "Academic Misconduct."

The Ohio State University's Code of Student Conduct (Section 3335-23-04) defines academic misconduct as: "Any activity that tends to compromise the academic integrity of the University, or subvert the educational process." Examples of academic misconduct include (but are not limited to) plagiarism, collusion (unauthorized collaboration), copying the work of another student, and possession of unauthorized materials during an examination. Ignorance of the University's Code of Student Conduct is never considered an "excuse" for academic misconduct, so I recommend that you review the Code of Student Conduct and, specifically, the sections dealing with academic misconduct.

If I suspect that a student has committed academic misconduct in this course, I am obligated by University Rules to report my suspicions to COAM. If COAM determines that you have violated the University's Code of Student Conduct, the sanctions for the misconduct could include a failing grade in this course and suspension or dismissal from the University.

If you have any questions about the above policy or what constitutes academic misconduct in this course, please contact me. Other sources of information on academic misconduct (integrity) to which you can refer include:

- The Committee on Academic Misconduct web pages (COAM Home)
- Ten Suggestions for Preserving Academic Integrity (Ten Suggestions)
- Eight Cardinal Rules of Academic Integrity (www.northwestern.edu/uacc/8cards.html)

Proposed Lecture Schedule

Date	Topic	Reading	HW
Jan 6th & 8th	Introduction to Spatial Data Analysis	Chi & Zhu, Chapter 1; Goodchild, 2000; Anselin, 1989	
Jan 13th & 15th	Correlation & Global Spatial Autocorrelation	Chi & Zhu, Chapter 2.1 - 2.4.3	
Jan 20 & Jan 22nd	NO CLASS 1/20 (MLK Day) Local Spatial Autocorrelation	Chi & Zhu, Chapter 2.4.4	
Jan 27th & 29th	Local Spatial Autocorrelation Bivariate Regression	Hair, et al, pp 151-161	HW 1
Feb 3rd & 5th	Multiple Regression (I)	Hair, et al, pp 161-181	
Feb 10th & 12th	Multiple Regression (II)	Hair, et al, pp 182-203	
Feb 17th & 19th	Regression Diagnostics	Hair, et al, pp 203-230	HW 2
Feb 24th & 26th	Regression Diagnostics (Spatial)	Chi & Zhu, Chapter 3.1	
March 2nd & 4th	Generalized Linear Regression (I)	Gelman & Hill, pp 79-108	Exam 1
March 9th & 11th	NO CLASS – SPRING BREAK	-	
March 16th & 18th	Generalized Linear Regression (II)	Gelman & Hill, pp 109-117	HW 3
March 23rd & 25th	Spatial Regression (SAR)	Chi & Zhu, Chapter 3.2	
March 30th & April 1st	Spatial Regression (SEM)	Chi & Zhu, Chapter 3.3	
April 6th & 8th	Spatial Regression (SEMSLR)	Chi & Zhu, Chapter 4.1 - 4.2	
April 13th & 15th	Geographically Weighted Regression	Chi & Zhu, Chapter 5.1, 5.3 Fotheringham & Brunsdon, Chapter 1	HW 4
April 20th	Catch-up and Conclusion	-	Exam 2