PLSC 510/510H/640
SOCIAL SCIENCE DATA ANALYSIS (FALL 2018)

Dr. Hye-Sung Kim
Office: Bancroft 326
Office Hours: Tuesdays 9:00-11:00 am; 12:30-2:30 pm
Email: kimh@winthrop.edu

Course Information
Section 001, 3 Credits
Classroom: Owens G04
Class Time: 5:00 pm - 6:15 pm

COURSE DESCRIPTION

This course is an introduction to the theory and application of linear modeling to social science problems. The focus of this class will be to provide you with the theoretical and practical skills necessary to conduct your own empirical research. Topics will include ordinary least squares, hypothesis testing, dealing with violations of the underlying assumptions of multiple regression, instrumental variables estimation, simultaneous equations, time-series econometrics, panel data techniques and causal inference, experiments, quasi-experiments and maximum likelihood estimation.

This course, however, differs from a standard research method or data analysis course because it is designed to equip students with skills that they will be able to apply to actual datasets to write publishable research papers and to apply for data science jobs after they complete the courses. For this purpose in mind, I will work extensively with each student individually to meet each student's needs as their level of familiarity with materials, previous experience, and the goals after taking the course are different. There are three routes that each student can take:

1. *Understanding Data Analysis and its Application to Solve Social Science Problems:* this is ideal for students who are enrolled in PLSC 510 and who do not have an in-depth statistical background. Students who desire to have employable data science skills in the data science industry and be prepared for more advanced training after the course such as attending data science camp are encouraged to take this module.

2. *Applying Data Science Techniques for Conducting An In-depth Social Science Research:* this is ideal for students who are enrolled in PLSC 510H and PLSC 640, who have already taken PLSC 350 or equivalent, Honor's students who are working on their Honor's thesis who uses the quantitative methods, MLA students with a Social Science question that requires to use quantitative methods to address the research question.

By the end of the course, students should be able to:

Identify the most appropriate methodological techniques for analysis given a research question and available data, as well as identify, understand the implications, and offer resolution to various problems encountered during quantitative analysis.

Conduct data analyses using the methodologies covered in the course. In particular, students should be able to diagnose and test empirical models and apply the techniques for correcting models that violate statistical assumptions.

Write a research paper using the data analysis tools rigorously we learned in class to address Social Science research questions, present the results at major conferences in the discipline, and submit their paper to a journal for publication.

Manage data and conduct analyses using R and RStudtio.

This course also contributes to the mastery of the following university-level competencies (ULCs):

1. Winthrop graduates think critically and solve problems.

Winthrop University graduates reason logically, evaluate and use evidence, and solve problems. They seek out and assess relevant information from multiple viewpoints to form well-reasoned conclusions. Winthrop graduates consider the full context and consequences of their decisions and continually reexamine their own critical thinking process, including the strengths and weaknesses of their arguments.

3. Winthrop graduates understand the interconnected nature of the world and the time in which they live.

Winthrop University graduates comprehend the historical, social, and global contexts of their disciplines and their lives. They also recognize how their chosen area of study is inextricably linked to other fields.

4. Winthrop University graduates communicate effectively.

Winthrop graduates communicate in a manner appropriate to the subject, occasion, and audience. They create texts – including but not limited to written, oral, and visual presentations – that convey content effectively. Mindful of their voice and the impact of their communication, Winthrop graduates successfully express and exchange ideas.

WHO SHOULD AND SHOULD NOT TAKE THIS COURSE?

Though PLSC 350 or equivalent is required, students who are taking PLSC 350 or equivalent concurrently can be granted to enroll in this course for exceptional cases.

REQUIRED TEXTS

- Bailey, M. (2015). *Real Stats: Using Econometrics for Political Science and Public Policy*. Oxford University Press.

- Free Online R Tutorial: https://cran.r-project.org/doc/contrib/Paradis-rdebuts_en.pdf

- Springer's Free R Resources **Use R!** https://link.springer.com/content/pdf/10.1007%2F978-0-387-93837-0.pdf

RECOMMENDED TEXTS

- Imai, Koske. (2017). *Quantitative Social Science: An Introduction.* Princeton University Press, 2017).

- Wooldridge, J. M. (2015). *Introductory Econometrics: A Modern Approach*. Nelson Education. (for beginners and intermediate level students)

- Fox, J. (2015). *Applied regression analysis and generalized linear models*. Sage Publications. (for advanced level students)

- Kennedy, P. (2008). *A guide to econometrics*. Wiley-Blackwell. (for advanced level students)

- Greene, W. (2010) *Econometric Analysis.* Pearson. (for advanced level students)

- Angrist, J. D., & Pischke, J. S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press. (for all levels)

R AND R STUDIO

For this course, you are required to use R and RStudio. In Owens 204 and other computer labs on campus, they have already been installed for your use. Both R and RStudio are free open-source software, downloadable free to your personal computer. To install R and RStudio, following these steps:

1. Download R first from the following website: https://cran.r-project.org/. There are three options, which will look like the following. Download and install one that matches your PC's operating system.

   - Download R for Linux
   - Download R for (Mac) OS X
   - Download R for Windows

2. Download R Studio from the following website: https://www.rstudio.com/products/rstudio/download/. You can choose "R Studio Desktop Open Source License FREE." Once you choose the link, you will be asked to choose a download option depending on your PC's operating system. Choose the version that matches your operating system.

   RStudio 1.1.456 - Windows Vista/7/8/10   85.8 MB   2018-07-19

   RStudio 1.1.456 - Mac OS X 10.6+ (64-bit) 74.5 MB   2018-07-19

COURSE REQUIREMENTS

- **Two Data Analysis Essays (20 % for PLSC 510 Students; 10% for PLSC 510H and PLSC 640 Students)**

  There will be two data analysis essay worth 20 % of your course grade. Students will be given a social science dataset and a research question to examine. In the essay, students will propose the best research methods to examine the given question and dataset, conduct a data analysis, and write the interpretations and results of the analysis. Each essay exam is worth 10%.

- **Problem Sets and Lab Report (20 % for PLSC 510 Students; 10% for PLSC 510H and PLSC 640 Students)**

  You will be assigned 8 problem sets, which will count for 20% of the course grade. Each problem set will require you to work on the given dataset and provide an R-script as well as answers to the questions in the problem sets. In addition, students are supposed to write-up "lab report" part where they write in detail about what statistical concepts, theorems, the R-code they used in answering each question and why they thought they were appropriate to use. If students encountered errors, also report them and whether they were able to figure out the problem, what was the cause of an error, or what remained unresolved. So, the purpose of this problem sets is to give you ample opportunities to practice how to conduct data analysis rather than getting the answers and code right. Your grade will not so much reflect on how much error-free your assignment is but more how much progress your entire assignment shows in the process of conducting a data analysis, and how much critical thinking you show in solving problems, trouble-shooting error messages in R and etc.

  You are encouraged to work in groups (composed of no more than 3 students). Since a solution set will be available on the course website as soon as a problem set is handed in, late problem sets will not be accepted (which means you will receive a zero).

- **Group Research Project (60 % for PLSC 510 Students; 30% for PLSC 510H and PLSC 640 Students)**

  Students can sign up for one of the following research group.

  1. *Do black lives matter equally?* The Washington Post's recent article showed the rate for black homicides was solved at much lower rates than white homicides. In some cities, the homicides of black victims are solved at half the rates of white homicides. What explains this discrepancy? The police say it is the location – black victims tend to live in places where it is harder to get witnesses and say they "don't care what color you are."

  2. D*oes economic growth decrease the likelihood of civil conflicts?* Using an instrumental variable approach and the data of 41 African countries during 1981–99, Miguel, Satyanath and Sergenti (2004) found that economic growth and the likelihood of civil conflicts are negatively related: negative economic shocks increase the likelihood of civil wars and positive shocks decrease it. In this project, will update the data used in the Miguel, Satyanath and Sergenti (2004) and replicate their analysis.

     Reference: Miguel, E., Satyanath, S. and Sergenti, E., 2004. Economic shocks and civil conflict: An instrumental variables approach. *Journal of Political Economy*, *112*(4), pp.725-753.

  This requirement consists of the following components:

  1. A critique note of the main paper to be replicated, consisting of the research question, data sources, the research methods, the plan on how replication will differ from the original paper, and the list of references (5% for PLSC 510; half the worth for PLSC 510H/PLSC640 Students)

  2. Annotated bibliography (10% for PLSC 510; half the worth for PLSC 510H/PLSC640 Students)

  3. Complete Dataset Construction (10% for PLSC 510; half the worth for PLSC 510H/PLSC640 Students)

  4. Results and Data Analysis (10% for PLSC 510; half the worth for PLSC 510H/PLSC640 Students)

  5. First Draft (5% for PLSC 510; half the worth for PLSC 510H/PLSC640 Students)

  6. Final Paper (10% for PLSC 510; half the worth for PLSC 510H/PLSC640 Students)

  7. Abstract submission to 2019 Midwest Political Science Association Annual Meeting and 2019 SOURCE conference (Due TBA); (5% for PLSC 510; half the worth for PLSC 510H/PLSC640 Students)

  8. Presentation (5% for PLSC 510; half the worth for PLSC 510H/PLSC640 Students)

     On November 28, all students will present their original research papers to the class. Each student will be allotted 12-15 minutes for presenting and each presentation will be followed by 3-5 minutes' worth of questions from the class. Each student will present the following about their assigned paper: the research question, the data used, the empirical methodology, and the findings. The presentation will include an interpretation of the findings and how these results answer the original research question. If your paper is not turned in by the beginning of the day of presentations, you will not be able to present and you will receive a zero for the presentation. No exceptions.

- **Original Research Paper (Not required for PLSC 510 Students; 50 % for PLSC 510H and PLSC 640 Students)**

This paper will count for 50 % of the course grade. Conduct empirical research on the question of your choice using a dataset of your choice. A research proposal that includes the dataset, the research question and details about the key variables of interest. This paper will include a brief introduction to the research question, a literature review, a description of the data and measures, including descriptive statistics, an explanation of the empirical methodology, results, and the discussion/interpretation of the results. The style of the paper should be similar to that of published journal articles. Unless you receive prior authorization from the instructor in writing, late papers will be penalized two full letter grades for each day---or fraction thereof that they are late. This paper consists of the following components:

1. Research proposal consists of the research question, data sources, and the list of references (4%)
2. Abstract submission to 2019 Southern Political Science Annual Meeting (Due September 1, 2018) and/or 2019 Midwest Political Science Association Annual Meeting (Due TBA); (4%)
3. Evaluation of five potential journals that you find the best fit for the research topic so that you may submit your manuscript to (4%)
4. Annotated bibliography (5%)
5. Research Design (5%)
6. Complete Dataset Construction (5%)
7. Results and Data Analysis (5%)
8. First Draft (4%)
9. Second Draft (4%)
10. Final Draft (5%)
11. Presentation (5%)

    On November 28, all students will present their original research papers to the class. Each student will be allotted 12-15 minutes for presenting and each presentation will be followed by 3-5 minutes' worth of questions from the class. Each student will present the following about their assigned paper: the research question, the data used, the empirical methodology, and the findings. The presentation will include an interpretation of the findings and how these results answer the original research question. If your paper is not turned in by the beginning of the day of presentations, you will not be able to present and you will receive a zero for the presentation. No exceptions.

- **(Optional) Winthrop Poll**

  Participating in the Winthrop Poll as a caller or supervisor will provide students valuable skills as a researcher especially those who are interested in conducting a survey research. Completing one shift successfully can earn up to 5% of extra credits toward the final course grade.

GRADING

Course Grades will be distributed as follows:

96 or above	A
93-95	A-
87-89	B+
83-86	B
80-82	B-
77-79	C+
73-76	C

70-72	C-
67-69	D+
63-66	D
60-62	D-
Below 60	F

The "N" Grade This semester, the deadline to withdraw from a course with an automatic grade of "N" is October 19th. Students may not withdraw from the course after this date without documented extenuating circumstances.

ACADEMIC INTEGRITY

All students are responsible for maintaining the highest standards of honesty and integrity in every phase of their academic careers. The penalties for academic dishonesty are severe and ignorance is not an acceptable defense.

TURNITIN

ALL written assignments must be submitted to Turnitin.com by the due date in order to be graded. The Class ID is 18788219 and the Enrollment Key is dataanalysis.

COURSE TOPICS
Note: The course syllabus is a general plan for the course; deviations announced to the class by the instructor may be necessary. We may not cover all of these topics. Conversely, time permitting, other topics might be covered in this course.

Dates	Class Topics	Assignments and Deadlines
August 20 and August 22	**Topic 1** Introduction to Social Science Data Analysis & Introduction to R and RStudio Bailey, Chapter 1	**ALL** • Data Assignment/Problemt Set 1: R & RStudio installation: August 22
August 27 and August 29	**Topic 1 (cont'd).** Introduction to Data Analysis & Introduction to R and RStudio	**PLSC 510H/PLSC 640 Students** • Research Proposal: August 27 • An Abstract To be Submitted to SPSA (August 29)
September 3: ***Labor Day (No Class)*** September 5	**Topic 2: Introduction to Simple Regression** Bailey, Chapter 3 • OLS estimator as random variables • Six assumptions of the classical linear regression model • The Gauss Markov Theorem • Unbiasedness; Precision of Estimates; Consistency • The R-squared goodness of fit measure	**ALL** • Data Assignment/Problem Set 2 (September 5)
September 10 and 12	**Topic 3.** Interval estimation and hypothesis testing	**ALL**

	Bailey, Chapter 4 • Common families of statistical distribution • OLS under the normality assumption • Confidence intervals/interval estimation • Testing hypothesis about a single parameter; the t-test and statistical significance • Difference between statistical significance and economic (substantive) significance • Testing hypothesis involving several parameters: F-test	• A Critique Note of Group Project Model Paper (September 12) **PLSC 510H/PLSC 640 Students** • Journal Evaluation Due (September 12)
September 17 and 19	**Topic 4** Multiple linear regression Bailey, Chapter 5 • OLS and goodness of fit in the K-variable model • The classical assumptions revisited • Endogeneity • Precision and goodness of fit	**ALL** • Data Assignment/Problem Set 3 (September 19)
September 24; September 26	Topics 1 through 4 Review In Class Data Analysis #1 Exam	**ALL** • Annotated Bibliography (September 24) • Data Analysis Exam #1: In Class (September 26) **PLSC 510H/PLSC 640 Students** • Research Design (September 26)
October 1 and 3	**Topic 5** Model specification and functional form Bailey, Chapters 5.5; 6; 7 • Model specification • Functional transformations of dependent and independent variables • Functional form test • Dichotomous (dummy) independent variables • Interaction variables	**ALL:** Interim Grade Deadline (October 4 5:00 pm) • Data Assignment/Problem Set 4 (October 1) • Complete Dataset of Group Project (October 3)
October 8 and 10	**Topic 6** Model specification, omitted variable bias and other data problems Bailey, Chapters 5.2; 5.3; 5.5; 14.1 through 14.5; • Omitted variable bias • Omitted variable bias tests	**ALL:** • Data Assignment/Problem Set 5 (October 8) **PLSC 510H/PLSC 640 Students**

	• Including irrelevant variables • Comparing nested and non-nested models • Different effects across groups • Multicolinearity • Measurement errors • Heterocedasticity • Diagnosis and remedy of heteroscedasticity • Autocorrelation • Diagnosis and remedy of autocorrelation	• Complete Dataset for Research Paper (October 10)
October 15 *Fall Break (No Class)* October 17 and October 22	**Topic 8** Instrumental Variables Estimation and Two Stage Least Squares Bailey, Chapter 10; Miguel, E., Satyanath, S. and Sergenti, E., 2004. Economic shocks and civil conflict: An instrumental variables approach. *Journal of political Economy*, *112*(4), pp.725-753. • Experiments • Instrumental variables • 2SLS	**ALL:** • Data Assignment/Problem Set 6 (October 22)
October 24	**Topics 9** Time-series econometrics Bailey, Chapters 13 • Modeling autocorrelation • Detecting autocorrelation • Fixing autocorrelation • Dynamic models • Stationarity	**PLSC 510 Students** • Results and Interpretation of the Group Research Project (October 24)
29 and 31	**Topics 10** Panel-data econometrics Bailey, Chapters 8; 15 • Difference-in-difference • Fixed effects models • Random effects models	**ALL:** • Data Assignment/Problem Set 7 (October 29) **PLSC 510H/PLSC 640 Students** • Results and Data Analysis for Research Paper (October 31)
5 and 7	**Topics 11** Regression Discontinuity Bailey, Chapter 11 • Basic RD Model • More Flexible RD Models • Windows and Bins • Limitation and Diagnostics	**PLSC 510 Students** • First Draft (November 5) **PLSC 510H/PLSC 640 Students** • First Draft of Research Paper (November 7)
12 and 14	**Topic 12** Logit and Probit Models for Categorical Response Variables: Models for Dichotomous Data Bailey, Chapter 12; Fox, Chapter 14.1 • Using latent variables to explain observed variables	**ALL:** • Data Assignment/Problem Set 8 (November 12)

	• Probit and logit models • Maximum likelihood estimation • Interpreting probit and logit coefficients	
November 19 **November 21: Thanksgiving Break: (No Class)**	Topics 5-12 Review	**ALL** • Abstract Submission for MPSA and SOURCE for Group Project (November 19) **PLSC 510H/PLSC 640 Students** • Abstract Submission for MPSA and SOURCE (November 19) for Research Paper • Second Draft (November 19)
November 26	In Class Data Analysis #2 Exam	Data Analysis Exam #2: In Class (November 26) **PLSC 510 Students** • Final Paper of Group Project (November 26) **PLSC 510H/PLSC 640 Students** • Final Paper (November 26)
November 28	Paper presentations	**ALL:** • GADGET DAY & PAPER PRESENTATIONS

UNIVERSITY LEVEL CLASS POLICIES AND RESOURCES

Attendance, Class Participation, and Taking Notes: Regular class attendance is required. Excessive absences will lower your grade in two ways: (1) loss of note-taking from class lecture and discussion; (2) loss of class participation credit. Class participation will raise or lower your grade. If prolonged illness or other problems cause you to be absent for an extended period of time, please let me know. In addition, every student is expected to arrive on time. Arriving late or leaving early disrupts the class and is not acceptable.

Class Listserv Participation: Class cancellations, changes in schedule, relevant global cultural events, guest speakers, etc., will be announced via the class listserv to which you will be automatically subscribed through your winthrop.edu e-mail address if you are registered for the class by the time the listserv population is generated. The email address for this list is: PLSC510001@class.winthrop.edu. If you register later and need to subscribe to the class listserv, go to:

http://www.winthrop.edu/technology/default.aspx?id=7081. The class number is **PLSC510001**. Check your WU e-mail frequently. You may use the listserv for discussion or to share information with classmates, review for exams, or ask questions about course material. *Note that anything you post to the listserv will be seen by everyone in the class.*

Technology: You are welcome to use a computer to take notes in class. However, using the computer to check Facebook or engage in other non-course related activities is prohibited. This class follows the College of Arts and Sciences policy on use of technology in the classroom. Please turn off all cell phones and other electronic devices during class meetings. If you need to leave your phone on during a particular class for emergency reasons, please place it on vibrate. *The use of any kind of electronic device, including a phone, iPod or Internet access, during an exam constitutes academic dishonesty.*

http://www.winthrop.edu/uploadedFiles/artscience/AppropriateUseOfHandHeldWirelessTechnology ApprovedPolicyMar2010.pdf

Office of Accessibility (OA): Winthrop University is dedicated to providing access to education. If you have a disability and require specific accommodations to complete this course, contact the Office of Accessibility (OA) at 323-3290 or accessibility@winthrop.edu. Once you have your official notice of accommodations from the Office of Accessibility, please inform me as early as possible in the semester.

Winthrop's Academic Success Center is a free resource for all undergraduate students seeking to perform their best academically. The ASC offers a variety of personalized and structured resources that help students achieve academic excellence, such as tutoring, academic skill development (test taking strategies, time management counseling, and study techniques), group and individual study spaces, and academic coaching. The ASC is located on the first floor of Dinkins, Suite 106. Please contact the ASC at 803-323-3929 or success@winthrop.edu or www.winthrop.edu/success.

Winthrop University's Office of Nationally Competitive Awards (ONCA) identifies and assists highly motivated and talented students to apply for nationally and internationally competitive awards, scholarships, fellowships, and unique opportunities both at home and abroad. ONCA gathers and disseminates award information and deadlines across the campus community, and serves as a resource for students, faculty, and staff throughout the nationally competitive award nomination and application process. ONCA is located in Dinkins 222B. Please fill out an online information form at the bottom of the ONCA webpage www.winthrop.edu/onca and email onca@winthrop.edu for more information.

Plagiarism Using the words or ideas of others as one's own is plagiarism. Quoting or paraphrasing material from books or articles without properly citing the source is also plagiarism. All sources used must be properly cited in your papers. Consult your Writing 101/HMXP Writing Manual for proper citation techniques.

POLITICAL SCIENCE DEPARTMENT STATEMENT ON PLAGIARISM AND ACADEMIC MISCONDUCT

The Winthrop University Political Science department abhors all forms of academic misconduct, and faculty members aggressively investigate all incidents of suspected cheating. This includes, but is not limited to, using turnitin.com. Plagiarism, whether intentional or unintentional, is by far the most common form of academic misconduct in the department. Plagiarism includes, but is not limited to:

- Using the words or ideas of others as one's own;
- Reproducing, in whole or in part, principal ideas from a fellow student's work;

- Granting a fellow student permission to copy one's paper, or to reproduce some or all of its principal ideas;

- Quoting or paraphrasing material from sources without any citation;

- Quoting or paraphrasing material without sufficient and/or proper citation;

- Omitting some or all sources used in a paper; and

- Submitting a paper written for one course -- whether in Political Science or another discipline -- to meet a course requirement in a second course, *without the express permission of all instructors involved.* This is the case even though many paper topics may be relevant to several different courses.

All incidents of suspected academic misconduct are investigated with equal vigor. When a faculty member suspects that a student engaged in academic misconduct, the faculty member will follow the appropriate procedures outlined in the *Student Handbook*. The faculty member will apply whatever sanctions s/he deems appropriate. Possible sanctions include, but are not limited to:

- Failing the assignment;

- Requiring a student to repeat an assignment for reduced credit;

- Requiring a student to repeat an assignment for no credit; or

- Failing the course.

Academic misconduct applies equally to required assignments and extra credit assignments.

All incidents of academic misconduct will be reported to the Department Chair, the Dean of Students, the Dean of Arts and Sciences and the student's academic advisor. The University may impose its own sanctions in addition to sanctions imposed by the faculty member or the department. The University may impose sanctions even after a student has graduated, and may include revoking a student's diploma.

In addition, students who engage in more than one incident of academic misconduct may be declared ineligible for departmental awards, ineligible for employment in the department or its affiliated programs, and ineligible to volunteer as a peer advisor.

Adopted August 14, 2007.