**Sociology 361, Statistics for Sociologists II, Fall, 2012**
Sewell Social Science Bldg 6102, TuTh 9:30AM - 10:45AM

**Prof. John Allen Logan**
email logan@ssc.wisc.edu
phone 262-0995
office hours W 10-12, and by appt.
4438 Sewell Social Science Bldg.

**TA Bo Hee Min**
email bohee.min@gmail.com
phone TBA
office hours Th 1-3, and by appt.
7110 Sewell Social Science Bldg.

**Labs**
301: Mon 1:20-3:15pm SS 6109
302: Mon 3:30-5:25pm SS 6109

**Subject matter and objectives:** This is a second course in social statistics, reviewing certain topics in basic inference, and then developing the theory and practice of applied regression analysis. A previous course on the basics of inference, hypothesis testing, and confidence intervals is presumed. The theoretical material will be developed in parallel with many practical exercises using Stata and SAS, the two predominant statistical packages for professional sociological research. Students completing the course will be ready to conduct their own research using basic regression analysis, and be qualified to become beginning research assistants on faculty research projects. They will have a solid foundation for studying the many methods that are extensions of regression analysis.

**Prerequisites:** Sociology 360, "Statistics for Sociologists I," or equivalent. Junior standing. Competence in basic algebra.

**Required Text:** Rachel A. Gordon, *Applied Statistics for the Social and Health Sciences*, 1st ed., Routledge. 2012. ISBN 9780415875363. Comment: The later parts of this text will also be used for part of Sociology 362, as I will teach it next Spring. So hang on to it if you plan to take that course as well.

**Homework:** Students are required to complete all Review Exercises and Chapter Exercises found at the ends of the chapters for all *assigned parts* of the chapters. This is a substantial time commitment. Your homework assignments constitute the biggest part of your grade (see below).

**Undergraduate vs. graduate student requirements:** Undergraduate students are permitted to complete their homework assignments using their choice of *either* SAS or Stata. Graduate students are required to use *both* SAS and Stata for all assignments.

**Course Exercise.** In addition to the Review and Chapter exercises, the book contains a cumulative Course Exercise, with associated tasks that are found at the ends of most chapters. This exercise asks you to formulate and investigate your own research question of interest to you, using secondary data that you will locate with the methods described in the course. You are required to complete these exercises, and to obtain the instructor's approval for your selected research question. All students may opt to use

either SAS or Stata for the Course Exercise (rather than both SAS and Stata), unlike the requirement for the Review and Chapter exercises described above.

**Computer lab sessions:** Most or all of the weekly lab sessions will be held in a computer lab instead of the room listed in the course schedule. Your TA and staff members of the Social Science Computing Cooperative (SSCC) will provide instruction and consultation on the use of SAS and Stata for course assignments.

**Computing outside the lab:** SSCC computers with SAS and Stata are available for course assignment use via the web. You will be offered advice concerning how best to get access from your home or from other campus computer labs or wifi connections. We require all students to begin their computing work in the SSCC computer lab, however. This simplifies getting everyone a basic familiarity with the systems and programs.

**Work independently:** All assignments are to be based on computer code independently written and run by each individual student. In addition, each assignment's written analysis must be the independent work of the individual student. Evidence of borrowing or copying between or among students will be (heavily) penalized. However, students are welcome to consult with each other in general terms about strategies for completing the various assignments. You **must turn in your own computer output** with each assignment that requires the use of a computer.

**Lab attendance and procedures:** Rules for lab attendance and for the form that homework assignment submissions must take are made and enforced by your TA. Your TA will also set deadlines for submitting assignments and determine whether late assignments can be handed in, and for how much credit.

**Pop quizzes:** There will be occasional pop quizzes intended to encourage you to read the chapters before the lectures are presented. The quizzes will consist of one or two of the Review Questions (*not* the Review Exercises) found at the end of the chapter that will be the topic of the day. The selected Review Questions will be presented unchanged from what is found in the textbook, giving you a convenient way to prepare for the quizzes.

**Grading:** Most of your grade (60%) will be based on the homework assignments. Twenty percent will come from the Course Exercise. The final twenty percent will come from any pop quizzes during lecture or lab, and your classroom and lab participation as qualitatively assessed by your TA and professor. There will be no final or other exams.

**Office Visits:** If you are habitually shy or quiet in class or lab please visit your TA and/or your professor in office hours so we can get to know who you are and discuss any problems or issues you may have.

**Departmental Notice:** The Department of Sociology regularly conducts student evaluations of all teaching assistants near the end of the semester. Students who have more immediate comments, complaints or concerns about the teaching assistant may

report them either to Professor Mara Loveman, Associate Chair, or Professor James Montgomery, Chair, 8128 Social Science.

**Reading Assignments:** My intention is to cover the indicated chapters of the textbook, stopping at Chapter 14. To repeat this important point: students will be expected to have read the relevant textbook material before each lecture.

**Topics.** Here is a synopsis of the topics of the chapters we will cover:

Chapter 1. Examples of social and health sciences research using regression analysis.

Chapter 2. Planning a quantitative research project with existing data.

Chapter 3. Basic features of statistical packages and data documentation.

Chapter 4. Basics of writing programs with statistical packages. How to write and run batch programs in Stata and SAS to create and modify variables for an analysis data set.

Chapter 5, *sections 5.6 and 5.7 only.* Weighted statistics, creating a descriptive table. *Important*: you will be responsible for reviewing and mastering the other material in the chapter on your own, as needed.

*Bonus topic*: *Creating Continuous Variables With Additive Scales.* Since many social science data sets have few continuous variables, we consider how to construct them from categorical (e.g. Likert) items, and how to assess their reliabilities.These skills may be useful in your course exercises.

Chapter 6. Sample, population, and sampling distributions. Statistical inference, population, sample, and *sampling* distributions. Basics of the *t*, *F*, and chi-square distributions.

Chapter 7. Bivariate inferential statistics. Some of this material may be covered in lecture, depending on the pace of the course and the interest of the students and instructor. Otherwise, students are responsible for reviewing the material independently. Section 7.5 on weighted statistics will be covered in lecture.

Chapter 8. Basic concepts of bivariate regression. Population and sample models, OLS, standard errors, hypothesis testing and confidence intervals, rescaling for substantive interpretation, standardized and semi-standardized slope coefficients, reversing the predictor and outcome, effect size (in the Cohen sense).

Chapter 9. Basic concepts of multiple regression. Multiple regression model, tests, intervals; rescaling for easier interpretation; standardized coefficients; conducting multiple tests; general linear F-test; model F-test; partial F-test; R-squared and adjusted R-squared; information criteria (AIC and BIC).

Chapter 10. Dummy variables. Dummies for binary and multi-category explanatory variables. Testing differences in means for the groups indexed by the dummies by: (a) re-estimating with a different reference category; (b) using the partial F-test; and (c) testing a linear combination of coefficients (by using the Var(A + B) = Var(A) + Var(B) + 2Cov(A,B) formula). Model interpretations with more than one dummy-coded variable, and with a mix of dummy-coded and continuous variables.

Chapter 11. Interactions. Interactions between two dummy explanatory variables; conditional regression equations (i.e., given a value of one of the variables) and examining significance of effects in the same [via options (a), (b), and (c) listed in Chapter 7 heading]. Same topics for interactions of a dummy with a continuous variable. Chow test. Presenting results in fully interacted models by holding other variables constant at their means.

Chapter 12. Nonlinear relationships. Qualitative consideration of shapes of nonlinear relationships. Quadratic and logarithmic transformations. Measuring change in Y as: absolute change, factor [multiplicative] change, proportionate [relative] change, percentage change. Different combinations of response and explanatory variable scalings: lin-lin; log-lin; log-log; lin-log; lin-sq. Cautions on directly comparing R-squared across response scalings. Calculating the expected values of Y rather than log(Y) when the model was estimated in log scale. Evaluating dummy-variable coding as an alternative to a nonlinear transformation, with an F-test.

Chapter13. Indirect effects and omitted variable bias. Confounders and spurious relationships. Mediators. Using side-by-side model estimations in tables to look for the effects on other coefficients as confounders/mediators are entered. Calculating direct and indirect effects in a three-variable path model. Suppressors. Omitted variable bias as the result of a possible indirect effect via a variable not included in the model. Calculating the expected sign of an omitted variable bias by considering the likely signs of the effects on the indirect path through the omitted variable.

Chapter 14. Outliers, heteroskedasticity, and multicollinearity.

Outliers and influential cases. Examining data for the possible causes of outliers. Diagnostic measures: hat values; studentized and standardized residuals; Cook's distance; DFFIT; DFBETA. Rule-of-thumb cutoff values for those measures. Graphing the variables and relationships (box and scatter plots). Heteroskedasticity: possible reasons for it; effect on standard error (but not expectation); heteroskedasticity-consistent standard errors (Huber-White). Multicollinearity: effects of; diagnosing it via: (a) Variance Inflation Factors; (b) significant model F statistic but no significant individual coefficients; (c) increasing standard errors in models as more controls are added. Remedies for multicollinearity.