

SOCG206 · QUANTITATIVE METHODS II  
Winter 2013 · Monday and Wednesday · 9:00 AM -10:20 PM · SSB 101  
Professor Isaac Martin · iwmartin@ucsd.edu · SSB 469  
Graduate office hours by appointment

“No sociologist, not even a senior scholar, must be too self-important to undertake many tens of thousands of quite trivial calculations, sometimes for months at a time. One cannot delegate such work entirely to mechanical assistants if one wants to get something out of it--and what comes out of it is often precious little. But even this little will not come, unless one has some definite idea about the point of the calculations, and, while one is calculating, about the general scope of the results. Ideas normally come only after one has prepared the ground with very hard work.”

--Max Weber, “Science as a Vocation,” 1918.

The heart of graduate training in sociology, and of the ongoing process of learning that comprises a successful career in sociology, is the learning of methods.

“Methods” in sociology is a shorthand for a body of accepted knowledge about how best to do research on society. You will have to make hundreds of high-stakes decisions in the course of your own research. Methods comprise the accumulated wisdom of people who have been there before you about how to make those decisions. Methods include principles for deciding what to observe, procedures for recording those observations, and techniques for analyzing those observational data once they have been recorded. This course addresses only the last of these topics—techniques for analyzing data that have already been collected—and of such techniques, it addresses only a narrow subset, namely linear regression and various other regression-based methods for analyzing quantitative data. I emphasize this point because I want you to understand where the material in this course fits into the discipline and practice of sociology. It is a small but important corner of the map. I also emphasize this point because I want you to understand that this course will not substitute for comprehensive training in research methods, and therefore that you will not be done learning methods when you are done with this course. Indeed, I hope you will never be done learning methods: that would be like running out of curiosity.

The particular subject of this course includes a subset of what are sometimes called generalized linear models. I take, and I will urge you to take, a pragmatic view of such models: they are good for describing certain kinds of data, and under certain conditions we can use these summary descriptions to make some interesting statements about the social world. That is all. The use of these models does not require you to embrace any particular epistemology or philosophy of science any more than (say) your use of English requires you to subscribe to any particular philosophy of language.

The particular models we will cover in this course are the workhorses of quantitative analysis in sociology. All of these models are in common use. It is therefore important for a sociologist to have a working knowledge of these statistical techniques, to know what they are good for, and to know what they are *not* good for. Achieving this working knowledge will call on you to exercise all of your intellect. It will require working through enough of the mathematical foundations to grasp what you are actually asking the computer to calculate when you “run a regression”; it will require acquisition of a technical jargon that may sometimes feel like learning a second language; and it will require practice *doing* these methods, making mistakes, and doing them over again. By the end of the quarter you will not be fluent but you will be on your way.

I will assume that you have taken SOCG 205 or the equivalent, and that you have some familiarity with regression analysis. We will be using Stata, and many of our Wednesday classes will be held in the computer lab. There will be weekly homework assignments due in weeks 2 through 9 (each worth 8% of the course grade) and a final paper (worth the remaining 36%). I will discuss these assignments in class. The main required texts for this course are:

J. Scott Long. 1997. *Regression Models for Categorical and Limited Dependent Variables*. Sage Publications.

Douglas A. Luke. 2004. *Multilevel Modeling*. Sage Publications.

There will be additional required texts and handouts on some specialized topics. I will make them available as we progress.

Some other texts I recommend:

J. Scott Long and Jeremy Freese. 2006. *Regression Models for Categorical Dependent Variables Using Stata*. Stata Press. This is a very useful companion volume for our required course text.

Andrew Gelman and Jennifer Hill. 2007. *Data Analysis using Regression and Multilevel/Hierarchical Models*. Cambridge University Press. This is an outstanding textbook that I would assign if we were using R instead of Stata.

Peter Kennedy. 2003. *A Guide to Econometrics*. MIT Press. "Econometrics" is what economists call applied statistics; there's nothing particularly economic about this book except the examples. This is a wonderful book for practitioners. My copy is well-thumbed.

Our schedule of topics is as follows, subject to modification on the fly if necessary:

#### Week 1. SOME RULES OF SOCIOLOGICAL METHOD

*Statistics are the servants of sociological method, not the masters: you may use them for many different purposes, but no statistical technique can substitute for a careful research design.*

#### Week 2. LINEAR REGRESSION

*Linear regression is a convenient way to describe many conditional distributions, and under certain very rare conditions it may even be a valid method for causal inference.*

Reading: Long, ch. 1-2.

#### Week 3. PRACTICING LINEAR REGRESSION

*Complications will arise when you run a regression; we will consider some of the more common ones, how to diagnose them, and what to do about them.*

Reading TBA

#### Week 4. REGRESSION WITH DICHOTOMOUS DEPENDENT VARIABLES

*When the outcome that we wish to explain is categorical then we must use a non-linear function to describe its conditional distribution, and do a bit more work to interpret the coefficients.*

Reading: Long, ch. 3

#### Week 5. REGRESSION WITH POLYCHOTOMOUS AND ORDINAL DEPENDENT VARIABLES

*We will consider common variations of logistic regression that are used (a) when the categorical dependent variable has more than two categories, and (b) when those categories are ordered.*

Reading: Long, ch. 5-6

#### Week 6. REGRESSION WITH COUNT DEPENDENT VARIABLES

*Counting things seems like the most straightforward kind of quantification; but when your dependent variable is a count of something, it presents another special case in which linear regression may not be appropriate.*

Reading: Long, ch. 8

#### Week 7. GOODNESS OF FIT

*It is easy to tell software to compute a statistical model that yields some sort of numerical output describing your data. But it is much harder to decide if the statistical model actually describes the data well. We'll consider what it means to describe data well, along with some statistical formulas that purport to measure "goodness of fit" and some conventional criteria for judging the output of those formulas.*

Reading: Long, ch. 4.

#### Week 8. MULTILEVEL MODELS

*When the conditional expectation function varies systematically among groups of some kind, we can often learn more about the outcome of interest by using a statistical model to describe that variation explicitly.*

Reading: Luke, ch. 1-2.

#### Week 9. TIME SERIES CROSS SECTION REGRESSION

*When our data take the form of repeated observations of multiple individuals, it turns out that we have a special case of a multilevel modeling problem.*

Reading: Luke, ch. 3, esp. pp. 62-72.

#### Week 10. EVENT HISTORY MODELS

*Sometimes we want to explain why events of a particular kind occur at a certain rate: for this purpose, we may use event history models that are closely related to many of the models we have already learned about.*

Reading TBA