SUBSTANTIVE AREA: **Methods and Statistics**
AUTHORED BY: **Bill Carbonaro**
DATE: **Last Revised September 8, 2016**

OVERALL SCOPE OF AND EXPECTATIONS FOR THE EXAM

Methods and Statistics are *huge* fields, ones that go far beyond just Sociology. Further, treatment of topics by different authors ranges from highly intuitive to highly mathematical. We do not expect students to know everything there is to know about statistics, or for all students to approach topics in the same way. Exams therefore usually provide some flexibility with regards to questions answered and how they are answered. But, we do expect students to be familiar with the basic methods and statistics that are widely employed in the Social Sciences and to have a solid understanding of several of the major advanced techniques that are popular in Sociology. Student answers should demonstrate that they have sufficient expertise to be a user of some advanced methods while being able to read, understand, and explain others. Solid mastery of the material in the required courses is assumed, but expertise well beyond that is expected.

Old exams and suggested readings are stored on the web at:

https://www3.nd.edu/~rwilliam/ndonly/area_exams/xsocmeth/index.html

While new topics, and new types of questions, will periodically be introduced, developing a mastery of old exam questions is an excellent way to prepare for new exams. Some (but not all) of the specific topics that have appeared in the past and are likely to appear again in the future include

- Basic techniques covered in the required methods and statistics classes. These include OLS regression, hypothesis-testing techniques, diagnostic procedures for problems such as heteroskedasticity and outliers, violations of assumptions and ways to test for and/or deal with them, dummy variable coding, interaction effects, nonrecursive models, nested models, standardized coefficients. “Simpler” methods such as t-tests and matched pairs studies should not be ignored, as students sometimes make basic mistakes on these.
- Test statistics and measures of fit, including Wald tests, likelihood ratio tests, BIC and AIC statistics.
- Categorical data analysis, including logit and probit models, loglinear analysis, count models, ordinal regression techniques (including both the ordered logit model and advanced techniques, such as heterogeneous choice models and generalized ordered logit models).
- Multilevel modeling (aka HLM), including the many different variants of random and fixed effects models. Important topics include: knowledge of both two and three-level models; random intercept and random slope models; cross-classified models; models analyzing longitudinal panel data; growth curve modeling with cross-level interactions; and applications of variance decomposition methods.

- Event history analysis, including: basic discrete-time hazard models; continuous-time event occurence models (e.g., continuous-time survivor and hazard functions); and Cox regressions models, and their extensions.
- Structural equation modeling. Applications of SEM incude: measurement models and scale construction; path analysis and cross-lagged panel models; dynamic panel models; instrumental variable estimation; and latent growth curve modeling.
- Adjusted Predictions; Marginal effects; and more generally, techniques for making results from models, especially nonlinear models, more tangible and easier to interpret.
- Causes and consequences of missing data, and strategies for dealing with it (both statistically after the data have been collected, and ways for avoiding it in the first place). Specifically, students must understand MCAR and MAR assumptions, and the limitations of conventional approaches such as listwise deletion and imputation. Knowledge of the application of multiple imputation and maximum likelihood methods for handling missing data is important, as well as the assumptions entailed with each approach.
- Techniques for cross-sectional panel studies. This includes both gathering the data and analyzing it once you have it. Other techniques for gathering data about the past, such as retrospective histories or successive cross-sections, should also be understood.
- The logic of causal ordering in statistical modeling. This includes an understanding of direct effects, indirect effects, common causes, spurious relationships, interaction effects, recursive and nonrecursive models
- The counterfactual framework for making causal inferences. This includes the theoretical foundations of the model, as well as statisical applications that attempt to identify causal effects using obseravation data. Students should have knowledge of directed acyclic graphs (DAG's) and their applications in identifying causal effects particularly in identifying which variables should (and should not) be conditioned on when making causal inferences.
- Propensity score matching. Students should have knowledge of the numerous different methods for generating propensity scores (e.g., logistic regression, optimized matching, etc.), and also different applications of propensity scores to balance observational data (e.g., matching techniques, stratification, and IPT weigthing). Students should also be familiar with different matching techniques that do not entail propensity scores (e.g., exact and coarsened matching), as well as techniques for testing the sensitivity of results to unobservables (e.g., Rosenbaum bounds).
- Survey design and data collection. This includes an understanding of methods such as stratified and cluster sampling, as well as an understanding of how the use of such techniques affects your statistical strategies.
- Experimental design. This includes a solid understanding of threats to internal and external validity and how different designs try (or fail) to deal with them. Students should also have a good understanding of the strengths and weaknesses of experimental studies compared to other methods of data collection, such as surveys.
- Assorted measurement issues, such as scale construction, reliability and validity, questionnaire construction, question wording, dealing with sensitive topics. Students You should be able demonstrate an understanding of how to collect better data in the first place, as well as understand how to statistically deal with measurement issues once the data have been collected.

- The ability to design a study on a topic of your choosing using a variety of research methods or statistical techniques, and/or the ability to explain how you would address a substantive topic that is given to you.
- Qualitative methods, including grounded theory, content analysis, participant observation, comparative historical analysis, unobtrusive measurement, ethnographic studies. Questions sometimes ask students to compare and contrast the strengths and weaknesses of qualitative versus quantitative methods.
- Ethical issues in research, including both historical cases that have informed human subjects protections, as well as practical issues involved designing studies with the goal of protecting human subjects.

In addition, an ability to give specific, concrete examples, or to briefly discuss how something could be estimated in Stata or some other program, will often help to make an answer clearer. In particular, given that students are often given the opportunity to design their own studies, they are strongly encouraged to spend time beforehand thinking through what such studies would entail. "Generic" answers that give general details about a method but that are unable to apply it to the specific questions asked are strongly looked down upon. While we do not generally ask for, or expect, a lot of equations, equations can sometimes be helpful for clarifying points.

We expect students to show a broad understanding of methods and statistics and to be thorough and precise in their answers. At the same time, we also expect students to be able to clearly communicate this knowledge to others. Thus, sometimes questions explicitly state that the intended audience is a beginning graduate student or a policy-maker who is not a social scientist. In general, it is good to write as though you are addressing an educated and intelligent audience, but one that is not necessarily familiar with the specific topic of the question.

Also, students are welcome and encouraged to let the committee know about their areas of expertise and the advanced courses they have taken. The committee is under no obligation to ask questions in those areas, but we will consider doing so.

EXAM PREPARATION

USE OF READING LISTS

The topics of methods and statistics are covered by multiple sources in multiple ways. Students may find some sources much easier to understand than are others. *Therefore, the committee does not require any specific set of readings.* However, the exam web page, given above, does include several suggested readings. Students do not necessarily need to go over those specific readings but they should be familiar with the topics covered in them. Students who prepare their own reading lists are encouraged to share them with the committee and have them posted on the web. Students are expected to be familiar with published work in statisical and methodological journals written for a sociological audience. In addition, students should review statistical and methodological papers in the top sociology journals (*ASR*, *AJS*, etc.) as well as pieces published in the *Annual Review of Sociology*.

REQUIRED/RECOMMENDED COURSES

The department requires that students take Soc 63997 (Linear Regression), Soc 73994 (Categorical Data Analysis), and Soc 63913 (Research Methods). While mastery of the material in these courses is essential, it is not sufficient for preparing for the exam. Students are encouraged to take advanced statistics courses such as Soc 73997 (Establishing Causal Inference) as well as other advanced research methods courses offered by the department (e.g. Soc 63957 – Historical and Comparative Sociology; Soc 63984 – Qualitative Analysis; Soc 63915 – Designing Qualitative Research). Students also often take statistics courses in Psychology on topics like structural equation modeling and multilevel modeling. Some students have taken workshops offered by ICPSR in Michigan or Scientific Software International in Chicago.

Students will likely find that taking advanced courses is a major help with the exam, but they will still need to do substantial additional outside reading. Students should allow adequate time for their preparation. Remember, students shouldn't be rushing to pass one exam; they should be preparing themselves for a career in which methods and statistics are among their areas of expertise, and such a career will require them to stay current with the literature in their fields.

EXAM FORMAT

The exam has separate sections for methods and statistics. Each half typically has at least two major essay questions. The exams also typically include a section on terms or short answers. The essays are designed to show depth in a specific area, while the terms or short answers show breadth across a variety of areas. Also, the essays often ask students to design a study of their own choosing, or else ask students how they would research a topic that is given to them.

The exam is closed book. Although it will not always be necessary, students should bring a scientific calculator to the statistics portion of the exam. We do not expect students to memorize complicated formulas, so if such formulas are deemed necessary they will be listed in the exam. However, students will be expected to recognize what those formulas are for and which ones are appropriate for the problem at hand. In most cases, any required calculations will be simple and straightforward; the challenge will be recognizing how to do the calculations in the first place.

EXAM SCHEDULING

Students will be asked to take their area exam sometime during the department's officially scheduled exam dates. If more than one student is taking the exam in a given semester, students will be asked to find a date/time that works for all examinees. The committee will write and offer ONLY ONE EXAM per semester. If students completely fail the exam, they must wait until the next semester before they can retake it. Students should notify the committee chair of their intention to take the exam a full semester before the exam will be scheduled.

EXPECTATIONS REGARDING CITATIONS

The committee prefers that students follow ASA guidelines regarding citations: author(s) and year of publication are preferable. Students need not memorize the titles of articles, but citing book titles is recommended. We would expect that most essays cite several sources, while one or two sources may be adequate for terms and short answers. Occasional references to introductory texts are acceptable but should not be heavily relied upon for the entire exam.

EXAM RESULTS

POLICIES REGARDING RE-WRITES

Students will be assigned one of three grades for their exam: pass with distinction, pass, or fail.

FEEDBACK

The committee will try to provide results within two to three weeks but faculty conflicts or exam scheduling (e.g. right before Christmas or during the summer) may result in longer turnaround times. After the committee renders its decision, the student will be notified of the results by e-mail. The student will then be asked to schedule a meeting with the committee chair to discuss both strengths and weaknesses of the students' performance. This meeting will be scheduled for all students, regardless of whether they pass or fail the exam. Other committee members may provide written feedback to each student, or they may set up a separate meeting to discuss the students' exam performance. This decision will be left at the discretion of faculty on the committee.