

Department of Sociology
University of North Carolina at Chapel Hill

ADVANCED METHODS
PH.D. Exam - Spring 2002
Friday 1 Mar 2002

Committee Kenneth Bollen, Guang Guo, François Nielsen (Chair)
Email francois_nielsen@unc.edu for questions.

Answer 4 questions out of 6.

Q1. Piece-wise exponential event history models and discrete-time event history analysis are two of the most popular ways of analyzing event history data. In both cases, the analyst needs to prepare the data by turning the original data into observation-time-unit data (for instance, turn individual data into individual-year or individual-month data). Typically, one estimates the piece-wise exponential model via a software procedure for Poisson regression and the discrete-time model by a logistic regression procedure. Discuss the following.

1. One often hears the statement that the standard errors of the parameter estimates from these models need to be upwardly adjusted because the input data have inflated the sample size. Do you agree with this statement? Explain your position.
2. Give the argument why the piece-wise exponential model can be estimated by the procedure for the Poisson regression or why the discrete-time model can be estimated by the logistic procedure (one or the other will be sufficient).

Q2. One feature of contemporary approaches to data analysis is closer attention to features of the data that may violate the assumptions of classical statistical theory on which model estimation and statistical inference are based. One area in which this more cautious perspective has been applied is the presence of outliers and influential cases. Discuss the following points in a multiple-regression models context (or another context of your choice that may be appropriate):

1. What are "outliers" and "influential observations", and in what way(s) do they represent a problem for data analysis?
2. What diagnostic tools are available to detect the presence of outliers and influential observations, and how appropriate are these methods given the presence of a single, versus several, outlying and/or influential observations?
3. Evaluate alternative remedial approaches to model estimation in the presence of outlying and/or influential observations.

Q3. Structural equations models (SEMs), or covariance structures models, have become a mainstay of empirical modeling in the social sciences. One event that was arguably crucial for the popularity of SEMs was the release of the program LISREL by Joreskog and Sorbom. Since the LISREL formalism was embodied in a practical estimation program, generations of researchers learned to conceptualize SEMs in the LISREL framework and notation. Since then, other approaches to SEMs have been proposed. Discuss the following issues:

1. What are the main advantages and drawbacks of the LISREL formalism in conceptualizing and estimating SEMs?

2. Are there alternative formalization approaches to LISREL, and what are their advantages and drawbacks?
3. Suppose you will be teaching a course on SEMs next year. Would you teach LISREL or another formalization? Why?

Q4. Researchers rely on a number of ways of interpreting the results from logistic regression and multinomial regression. Sociologists often interpret the sign of a regression coefficient and say that a certain variable has a positive or negative effect. Alternatively, one can use the odds ratio interpretation, conventional predicted probability interpretation, or simulated predicted probability interpretation. Discuss the advantages and disadvantages of each approach.

Q5. There is a literature on "sample selection", and there is a literature on "missing values". Discuss the relationship between these two literatures (or absence thereof) by examining the following points:

1. What are the main issues raised by the sample selection literature, by the missing values literature?
2. What are the main remedial approaches recommended in each of the two literatures?
3. In what ways are these issues and remedies related? In what ways are they not?

Q6. Is it possible to postulate a general statistical model that would incorporate all of the (main) quantitative methods of sociology? To answer this question, discuss the following points:

1. What are the primary methods that would have to be subsumed by such a general model?
2. If it were possible to define such a general model, what would the model have to do?
3. If not, why can't we do it?