RESEARCH METHODS AND STATISTICS AREA EXAM
September 2018

Committee: Bill Carbonaro (chair), David Hachen, Rich Williams

****

PART ONE: STATISTICS

****

SECTION I: Select **four** of the following eight items and briefly address the following for each one:

- Define or explain the main significance of the item; and
- Describe the item's importance for research practice.

1. Attenuation Bias
2. Latent Class Analysis
3. Consistent estimator ("consistency")
4. Lowess smoother
5. Random Effect
6. Autocorrelation
7. Average Marginal Effect
8. Brant Test

++

SECTION II: You **must** answer question two.

QUESTION 2

Answer the following questions regarding structural equation modeling (SEM).

I. Describe the model's key assumptions and concepts. Be sure to address the following: exogenous, endogenous, observed, latent variables; measurement, structural, recursive and non-recursive models.
II. Describe some applications of SEM as a method. Provide an example from an empirical paper that uses SEM to analyze data. What are the main advantages of this method, and what problems does it address that alternative methods cannot?
III. How does one assess the goodness of fit for SEM's? Discuss developments and debates regarding goodness of fit statistics and potential limitations of models evaluated as fitting the data well.
IV. Recent developments have led some methodologists to advocate using SEM to estimate random and fixed effects models, particularly when analyzing panel data. What are the main advantages of estimating random effects models within a SEM framework?

+++

SECTION III: Select and answer either question 3 or question 4.

QUESTION 3

I. Explain what "count variables" are, and why their use as dependent variables in an OLS regression can be problematic.
II. Discuss alternative modeling strategies that can be used instead. Explain the rationale for these strategies, the ways in which they are or are not superior to OLS, and how their parameters are interpreted.
III. Be sure to discuss such concepts as:
    - Poisson Statistics Regression
    - Overdispersion, and Negative Binomial Regression
    - Models for Truncated Counts
    - Hurdle Models
    - Anything else you think is especially important to understand.

QUESTION 4

Consider missing data due to "item non-response" in a cross sectional probability sample.

I. Consider a multiple regression model that relies on "listwise deletion" to handle missing data.
   A. How might the coefficients in this model be affected by this strategy (consider both efficiency and bias)?
   B. What assumption(s) are necessary to make listwise deletion a defensible strategy?

II. Describe statistical reasoning underlying the following alternatives to listwise deletion in a multiple regression analysis. Discuss the strengths and weaknesses of each approach, as well as the assumptions that are required for each.
   A. Mean substitution
   B. Multiple Imputation
   C. Full Information Maximum Likelihood (FIML)

Consider missing data in a longitudinal data set, where the same respondents are surveyed at five points in time.

III. Describe the distinctions between the following three types of missing data in this panel data set: 1) within-wave missing, 2) whole wave missing, and 3) missing due to attrition. What are the potential problems posed by each type of missing data for a longitudinal multiple regression model?

IV. How should an analyst address each of these three types of missing data in an longitudinal multiple regression analysis?

RESEARCH METHODS AND STATISTICS AREA EXAM
September 2018

Committee: Bill Carbonaro (chair), David Hachen, Rich Williams

==========
PART TWO: METHODS
==========

SECTION I: You **must** answer question one.

QUESTION ONE

In the absence of random assignment, it is notoriously difficult to make causal inferences in the social sciences. Select **three** of the five analytical designs below. Describe the method of analysis, and explain how they enable a researcher to argue that a relationship between a "treatment" variable and outcome is causal. For each design that you select, be sure to explain what describe what assumptions must hold for making a strong inference about causality.

1. Difference in differences
2. Regression discontinuity design (RDD)
3. Instrumental variable analysis
4. Propensity Score Matching
5. Fixed Effects Model

++

SECTION II: Select and answer **TWO** of the following THREE questions.

QUESTION TWO

The Catholic Church has been shocked once again by charges of sexual abuse by the clergy. Shaken, the Pope has commissioned Notre Dame, the greatest Catholic research institution in the world, to get to the bottom of this controversy. The Pope is determined to find out whether recent charges are only isolated incidents or merely the tip of an iceberg. He will spare no expense in getting to the truth.

There is general agreement that some sort of survey research is called for. Beyond that, confusion reigns. No one is quite sure who should be surveyed, let alone how. As one of the most talented and gifted students in the University, it is up to you to propose a coherent strategy to measure the prevalence of sexual abuse by church clergy.

In making your recommendations, you should address the following issues. (Remember, the Pope is not a social scientist, so you will need to define or explain many of the technical terms and concepts you are talking about.)

I. Who should the target population be? Clergy? Lay people? The young, the old? Why?
II. What sampling frames could be used to generate a probability sample of the population? Describe the strengths and weaknesses of different sampling frames that could be used. In particular, what sources of bias could threaten the validity of the findings?
III. Is a probability sample the best strategy, or would a nonprobability sampling strategy be better? Explain.
IV. How will construct your sample given that the study should focus on both historical patterns of the past (as far 50-60 years ago) , as well as contemporary (present-day) instances of abuses?
V. How can you get people to participate in your study, given the sensitive nature of the topic? What strategies can you get to ensure that respondents provide accurate response to your questions?

QUESTION THREE

Campbell and Stanley laid out the basics of experimental design more than 50 years ago. Now you are preparing to explain them to your students. Discuss the following:

I. Explain the main threats to internal validity that they described. Discuss **at least four** of the threats, providing brief examples of when each might occur. Briefly explain why "pre-experimental" designs (the one-shot case study, the one group pretest-posttest design, and the static group comparison) do such a poor job in controlling for these threats.
II. Now, present and discuss **two experimental** and **two quasi-experimental designs**, and explain how they try to control for threats to internal validity, and what (if any) threats may still remain.
III. Discuss the the following features of some experimental designs, and explain their importance when designing experiments::
    A. Counterbalancing
    B. Block design
    C. Experimental Realism
    D. Manipulation Check

QUESTION FOUR

In recent years, there have been several "scandals" or "controversies" in the social sciences that have been rooted in problems of research methodology. Some of these problems are due to poor research practices, while others are largely due to ethical transgressions. Some are a mix of both.

Discuss the following four scandals/controversies and explain the root problem of each one. Then, explain how better research practices and/or more rigorous ethical standards could prevent these problems from arising again in the future.

1. Accusations of misconduct and inaccuracies in Alice Goffman's *On the Run*.
2. The fraudulent research on gay marriage by Michael LaCour.
3. The "replication crisis" in psychology.
4. The failure of many polls (poll aggregators) to predict the outcome of the 2016 Presidential election.