RESEARCH METHODS AND STATISTICS AREA EXAM

March 2018

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

****

PART ONE: STATISTICS

****

SECTION I: Select **four** of the following seven 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.

a) Dynamic Panel Data Modeling using Maximum Likelihood
b) Average marginal effect (AME)
c) Robust standard error
d) Errors-in-variables regression
e) Lowess smoother
f) KHB (Karlson, Holm and Breen) method
g) Consistent estimator ("consistency")

++

SECTION II: Select and answer either question 2 or question 3.

QUESTION 2

Answer the following questions about the difference between random and fixed effects models.

A. Describe the key differences between random and fixed effects models. Be sure to highlight the different assumptions of each model. Also, please describe differences in the estimation and interpretation of the parameters in the two models.
B. Explain how the two models could be applied to the same data, and yet produce different substantive findings.
C. How should researchers decide whether a random or fixed effects model should be used to analyze his/her data?
D. Finally, describe the strengths and limitations of random and fixed effects models.

QUESTION 3

Categorical outcomes pose numerous statistical challenges for researchers. Answer the following questions regarding different modeling strategies for categorical dependent variables.

I. Consider an analysis in which the outcome variable is dichotomous: *voting* (whether a person voted in given election). A researcher wants to estimate how personal income (net of an array of controls) is related to the likelihood of voting. She knows that two commonly used models in this situation are the linear probability model and the logistic regression model.
   A. Compare how the linear probability model and logistic regression model differ in the parameter estimates that they produce.
   B. What are the advantages and disadvantages of using each approach, and how would you decide which model should be used?

II. Consider an analysis where the outcome variable is ordinal: *labor force participation (LFP)* (unemployed but searching, working part-time, or working full-time). A researcher is interested in examining how levels of education (net of an array of controls) are related to a person's LFP.
   A. Compare the strengths and weaknesses of three modeling strategies for this analysis: Ordinary Least Squares regression, multinomial logistic regression, and ordinal logistic regression.
   B. Describe the key assumptions regarding whether each model is appropriate for the analysis. Explain how you would deal with a violation of assumptions for each model.
   C. Explain which model you would likely select, and provide the reasons for your choice.

++

SECTION III: Select and answer either question 4 or question 5.

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?

QUESTION 5

Measurement error is a common problem that researchers must confront in their analyses. Consider the following two OLS regression models. Each model contains the variable "Happy" which is a subjective measure of a respondent's overall level happiness. This variable is an additive scale of five items, and it has an alpha reliability of .60.

$$Happy = \beta_0 + \beta_1(Income) + \beta_2(Age) + \beta_4(Female) + \mu \qquad (1)$$

$$Income = \beta_0 + \beta_1(Happy) + \beta_2(Age) + \beta_4(Female) + \mu \qquad (2)$$

I. Describe the difference between systematic and random measurement error in the variable "Happy."

II. Imagine that "Happy" suffers from random measurement error (RME).
   A. How would RME in "Happy" affect the coefficients and their standard errors in model (1)?
   B. How would the fit of the model be affected (relative to a model with no RME for "Happy"?)
   C. How would RME in "Happy" affect its slope and standard error in model (2)?
   D. Would RME in "Happy" in model 2 affect the slopes for "Age" and "Female" in model (2), if neither of these variables was measured with RME? Explain.
   E. For model (2), how might a researcher address the problem of RME in "Happy" in his/her analysis?

RESEARCH METHODS AND STATISTICS AREA EXAM

March 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. Interrupted time series design
5. Cross-lagged panel model

+++

SECTION II: Select and answer either question 2 or question 3.

<u>QUESTION TWO</u>

You are interested in researching: (1) the determinants of who consumes pornography, and (2) what effects it has on people's lives. There are many different strategies for studying these two questions. Below, four different possible approaches to addressing these two questions are presented.

1. Field research observing pornography addiction support groups
2. In person, open-ended interviews with a convenience sample of adults
3. Anonymous internet surveys of respondents self-reported behavior, and other demographic and personal information about themselves
4. Archival data of a random sample of people's internet browser histories (over the period of one week)

For each approach, explain what you see as the advantages and disadvantages of using this approach to answer research questions (1) and (2).

<u>QUESTION THREE</u>

The National Organization for Women is commissioning a study on changing attitudes of and towards women since NOW was founded in 1966. It is soliciting proposals for either:

A **qualitative analysis based on interviews** with dozens of women who were both active and not active in the feminist movement,

or,

A **content analysis based on published documents**.

Choose one of these strategies, and describe the following:

1) How the study would be conducted. Include information on how individuals or documents would be chosen, and how their verbal or written comments would be collected, coded and analyzed.
2) The strengths and weaknesses of the strategy.
3) The types of findings that might result and how they would be interpreted. You might give hypothetical examples of the kinds of statements you would receive or find and what these statements could imply about changing attitudes.

++
SECTION III: Select and answer either question 4 or question 5.

QUESTION FOUR

Gender discrimination is extremely difficult for researchers to document because a researcher must eliminate all other possible explanations of gender differences in a given outcome. Experimental designs are very well suited for eliminating threats to internal validity, and are often used to test for discrimination.

Shelley Correll and colleagues conducted a laboratory experiment to examine whether there is a penalty for “parenthood” for men and women in the hiring process. Here is their description of the experiment:

> “Paid undergraduate volunteers (84 men and 108 women) rated a pair of equally qualified, same-gender (either male or female), same race fictitious job applicants, presented as real, who differed on parental status. Race and gender was manipulated by altering first names on applicant files, and parental status was manipulated on the resume and human resource memo [that raters evaluated].”

Correll et al. found that mothers were less likely to be recommended for hiring, and at lower starting salaries than non-mothers. The authors found the opposite pattern for fathers and non-fathers.

1. Does this study offer compelling evidence that there is “motherhood penalty” in the labor market? Why or why not?
2. Evaluate the mundane realism and experimental realism of this study.
3. Do you think a manipulation check was necessary to bolster the claims of a “motherhood effect”? Explain.
4. Do you believe that reactive measurement effects (e.g., “demand characteristics” and “evaluation apprehension”) may have shaped the findings? Explain.
5. Evaluate the external validity of these findings.
6. Would this study be more convincing if it were conducted as a field experiment? Why or why not? Explain how you might redesign the study as a field experiment to make the findings more compelling.

QUESTION FIVE

Much early research about errors in surveys focused on the interviewer as a source of observational error. Summarize the major themes of research about the impact of the interviewer on measurement error. Your essay should review and evaluate the following:

1. The ways that the interviewer can contribute to error in survey data
2. Empirical findings about the impact of the interviewer on variable error and bias
3. Methods for reducing the interviewer's contribution to error
4. Recent research about the role of the interviewer in increasing the validity of data
5. The implications of research about the role of the interviewer on error in survey data for other observational methods, in which data are collected, such as participant observation.