# Final Exam Review (100 points total)

## Final Exam, Tuesday, March 19 10:15-12:15

The final exam is not explicitly cumulative. However, you will need to have a solid understanding of multiple regression models. You will have 2 hours to complete the test (please return by 12:15 via email). There are no restrictions on use of notes or books or other sources, but you will not likely have enough time to look up all of the answers. **You must complete this exam on your own**.

## Short Essay (40 points)

There will be 2 short 'essay' questions worth 20 points each. These are open-ended questions on definitions and <u>concepts</u>. Answers should be about 1 paragraph and *in your own words*. I will pick 2 of the following questions:

- Describe the two reasons why it is inappropriate to use OLS regression when the dependent variable is dichotomous? Explain two ways in which the regression model is modified in logistic regression to address these problems. What is the process used to fit the logistic regression and how does it differ from OLS regression?
- In words, define an odds ratio. Explain how to interpret odds ratios for positive and negative relationships, being sure to specify how an odds ratio is interpreted for both dichotomous and continuous independent variables. Illustrate these concepts with real or hypothetical examples not described in class or in the readings. What is the conceptual approach taken to quantify the total variance accounted for with logistic regression?
- Explain how the general approach to modification of the regression model used with logistic regression can be generalized for other types of regression models with non-continuous outcomes. Give two real or hypothetical examples not discussed in class or in the readings that would use two different types of alternative regression models (i.e., other than logistic for binary variables and OLS regression).
- What are the assumptions of OLS regression that make an alternative model necessary for a count dependent variable? What are the relevant assumptions when deciding between the two main types of count regression models (estimators) discussed in class? State which is the better choice between these two types in most circumstances and give the reasons why?

## **Multiple Choice (30 points)**

There will be 15 multiple choice questions worth 2 points each. These may be on <u>any of the assigned reading</u> or <u>the lecture material</u> from 2/20/24 (beginning with chi-square and simple logistic) through 3/14/24. The purpose of these questions is to make sure you have read the material and learned the concepts from the text and class lecture.

### **Computations and Interpretation of Results (30 points)**

There will be two computational or printout interpretation problems (15 pts each) that involve write-ups similar to those in Homeworks 2 and 3. These questions may include one or more of the following:

### Computations:

odds ratio from 2 X 2 contingency frequency table

### Interpret:

Regression diagnostics, 2 X 2 chi-square (including reporting appropriate percents from a contingency table), simple and multiple logistic regression (including odds ratios, confidence intervals for B or odds ratios, and significance tests of coefficients, the overall fit—likelihood ratio "chi-square" test, pseudo-R<sup>2</sup>), ordinal logistic and probit regression, and multivariate analysis of variance.