IDS 702 Midterm Topics - 2025

General

- The exam covers course material from the first day of class to Oct 7. In other words, every lecture except Model Selection. More detail on specific topics is provided below.
- The exam will be multiple choice and short answer. Last year's exam is provided to see the types of questions I will ask and the length of the exam. I will go over last year's exam in class on Tuesday, Oct 21st.
- I will not ask questions about R code syntax, though you may have to interpret code output, such as diagnostic plots or a model summary table.
- In general, you do not need to know formulas, though for conceptual understanding it may be helpful to understand the intuition behind a formula (e.g., that R^2 is 1 - the ratio of SSE to SST).
- For unit 1, I won't ask questions from the pre-class reading that I didn't cover in the lecture. For unit 2, you are responsible for the material covered in the pre-class videos.

Unit 1: Statistics Fundamentals

Introduction:

- population vs sample
- inference vs prediction

Probability:

- Know definitions/concepts, but formulas will be provided if needed for calculation

Probability Distributions:

- random variable, probability distribution, PDF/PMF, support, parameters, expected value

- Be able to identify a binomial and normal distribution
- You do NOT need to know the formulas for individual probability distributions (the fact sheets that you made in groups). Any needed PDFs/PMFs will be provided

Sampling Distributions/CLT:

- Define/describe what a sampling distribution is and the CLT
- Distinguish between standard deviation and standard error

Maximum likelihood estimation:

- Describe the overall idea (i.e., what is a statistical likelihood? What is the difference between a PDF and a likelihood?)
- Be able to go through the mathematical mechanics to calculate an MLE given a PDF (take the log, differentiate, solve for the parameter)
- Know that we want the following properties of estimators: unbiased, low SE, low MSE, consistency. You will not need to know or use the formulas for these.

Confidence Intervals & Bootstrap:

- Conceptually distinguish between the analytical and resampling methods of generating confidence intervals
- Interpret a confidence interval
- Understand the 3 factors that affect confidence intervals

Inference/Hypothesis Testing:

- Definitions: null and alternative hypotheses, p-value, significance level
- Conceptually distinguish between parametric and simulation-based inference
- Know when to use a two-sample test (i.e., within and across-sample independence)

Unit 2: Linear Regression

Intro to SLR/MLR:

- Distinguish between the theoretical model and the fitted model
- Least squares estimation (concept, not formula)
- Interpret estimates, p-values, and confidence intervals
- Matrix notation: identify dimensions of the MLR model components

Categorical Predictors & Interaction Terms:

- Interpret coefficient estimates for categorical and interaction terms
- write level-specific models
- write full and reduced models for a nested F test
- when to use an interaction term

Assessment & Assumptions:

- R^2 value
- linear regression assumptions
- connect the linearity, normality, and homoscedasticity assumptions to diagnostic plots
- Interpret plots to diagnose possible violations

Problems:

- Know conceptually leverage, Cook's distance, and VIF, though you won't need formulas
- Interpret diagnostic plot output

The model selection lecture (i.e., the Step Away from Stepwise reading) will not be on the exam.