IDS 702 Midterm Topics

Unit 1: Statistics Fundamentals

Study Design:

- population vs sample
- simple random sampling, cluster sampling, stratified sampling
- confounding variables

Probability:

- Know how to apply the definitions, but formulas will be provided

Probability Distributions:

- random variable, probability distribution, PDF/PMF, support, parameters, expected value
- 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 definition of bias and standard error. You will NOT have to calculate bias

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, non-parametric, and simulation-based inference
- When to use a two-sample t-test
- You do NOT need to know the other tests for which you made fact sheets

Unit 2: Linear Regression (9/24-10/3)

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