This review material provides you with a list of important theories and concepts we've learned from lectures before the first midterm. You can take it as a "table of content" to guide you to specific pages in lecture slides or section notes.

1 Simple Linear Regression

1.1 Introduction

- Correlation Versus Causation

- Experimental and Observational Data

- Empirical Problem

1.2 Population and sample regression models

- Population and sample

- Ordinary least squares (OLS)

- Formulas of $\hat{\beta}_0$ (fitted line goes through sample average) and $\hat{\beta}_1$ (ratio of covariance to variance), relationship between slope estimator and correlation coefficient, predicted values, and residuals

- Coefficient interpretation (**sign and size by now**; "holding everything else constant" when multiple linear regression)

- Probability Framework for Linear Regression (LR)

- 2 interpretations of Linear Regression (LR)

1.3 Assumptions for LR Model

- Assumption 1: $Y_i = \mathbf{X}_i'\beta + \epsilon_i$ and $\mathbb{E}\left[\epsilon_i | X_i\right] = 0$ (Key Concept 4.3.1 in Text, p.131)

- Assumption 2: $\{X_i, Y_i\}_{i=1}^n$ is an i.i.d. sample (Key Concept 4.3.2)

- Assumption 3: The matrix $\mathbb{E}\left[\mathbf{X}_i\mathbf{X}_i'\right]$ is invertible (non-singular) (Implicitly assumed in Text)

- Assumption 4: The matrix $\mathbb{E}\left[\mathbf{X}_i\epsilon_i^2\mathbf{X}_i'\right]$ is non-singular (Implicitly assumed in Text)

- Assumption 5: The random variables (X_i, Y_i) have finite fourth moments.

1.4 Properties of OLS estimator

- Under assumptions 1-3, $\mathbb{E}[\hat{\boldsymbol{\beta}}] = \boldsymbol{\beta}$, that is, $\hat{\boldsymbol{\beta}}$ is an unbiased estimator of $\boldsymbol{\beta}$.

- Under assumptions 1-3, $\hat{\boldsymbol{\beta}} \xrightarrow[n\to\infty]{p} \boldsymbol{\beta}$, and thus the OLS estimator $\hat{\boldsymbol{\beta}}$ is consistent for $\boldsymbol{\beta}$. In large samples, the OLS estimator should be close to β with high probability.

- Under all 5 assumptions, $\sqrt{n}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)$ will in large samples be approximately bivariate normal with mean vector equal to $[0,0]'$ and variance matrix $\hat{\mathbf{V}} = \left(\frac{1}{n}\sum_{i=1}^{n} \mathbf{X}_i\mathbf{X}_i'\right)^{-1} \left(\frac{1}{n}\sum_{i=1}^{n} \mathbf{X}_i e_i^2 \mathbf{X}_i'\right) \left(\frac{1}{n}\sum_{i=1}^{n} \mathbf{X}_i\mathbf{X}_i'\right)$ where $e_i = Y_i - \mathbf{X}_i'\hat{\boldsymbol{\beta}}$, i.e., $\sqrt{n}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) \xrightarrow[n\to\infty]{d} \mathcal{N}\left(\mathbf{0}, \hat{\mathbf{V}}\right)$.

1.5 Hypothesis Testing in Simple Regression

– Distribution of $\hat{\beta}$

– Definition of type I error, type II error, critical value, etc.

– Hypothesis testing for a single parameter - t Test

– Hypothesis testing for multiple parameters - Wald Test

– Hypothesis testing for multiple parameters - F-Statistic

1.6 Confidence Intervals in Simple Regression

- Confidence interval for a single parameter

- Confidence region for multiple parameters

1.7 Other Inference Related Concepts

- Conditional homoscedasticity

- Formula for the asymptotic variance under conditional homoscedasticity

- R-square

- Standard error of the regression

2 Multiple Regression

2.1 Motivation for Multiple Regression

- Interpret LR as Best Linear Predictor (BLP)

- Omitted variable bias: formula and assessment

2.2 Multiple regression

- Population model for multiple regression and interpretation of the parameters ("holding everything else constant")

- Formula for the OLS estimator (same as in simple regression model if written in matrix form)

- Assumptions (same as in simple regression model, only differ in the dimensions of the matrices)

- Properties of OLS estimator: consistency and asymptotic normality

2.3 Hypothesis Testing in Multiple Regression

- Distribution of $\hat{\boldsymbol{\beta}}$

- Hypothesis testing for a single parameter - t Test (same as in simple regression)

- Hypothesis testing for multiple parameters - Wald Test

2.4 Confidence Regions in Multiple Regression

- Confidence interval for a single parameter (same as in simple regression)
- Confidence region for multiple parameters

2.5 Other Inference Related Concepts

- R-squared and adjusted R-squared, and their interpretations

- Standard errors of the estimator in multiple regression under the additional assumption of conditional homoscedasticity

2.6 Nonlinear Regression Functions

- General non-linear regression model

- Polynomials

- Interaction terms: interactions between two binary variables, interactions between a continuous and a binary variable, interactions between two continuous variables

- Dependent variable transformations: log-linear, linear-log, and log-log specifications

- Application to empirical problem

2.7 Assessing Studies

- Internal validity: omitted variable bias, wrong functional form, measurement error, sample selection bias, simultaneous causality bias

- External validity