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
  - **Lecture Slides:** Handout1 *(pp.4-9)*
- Experimental and Observational Data
  - **Lecture Slides:** Handout1 *(pp.10-12)*
- Empirical Problem
  - **Lecture Slides:** Handout1 *(pp.13-17)*

1.2 Population and sample regression models

- Population and sample
  - **Lecture Slides:** Handout1 *(pp.18-20)*
  - **Section Notes:** Section3 *(pp.1-3)*
- Ordinary least squares (OLS)
  - **Lecture Slides:** Handout1 *(pp.21-33)*
- 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
  - **Lecture Slides:** Handout1 *(pp.33-35)*; Matrix form: Handout2 *(pp.14-16)*
  - **Section Notes:** Section2 *(pp.4-7)*
- Coefficient interpretation (**sign and size by now**; "holding everything else constant" when multiple linear regression)
  - **Lecture Slides:** Handout1 *(pp.36-39)*
- Probability Framework for Linear Regression (LR)
  - **Lecture Slides:** Handout2 *(pp.3)*
- 2 interpretations of Linear Regression (LR)
  - **Lecture Slides:** Handout2 *(pp.4-8)*

1.3 Assumptions for LR Model

- Assumption 1: $Y_i = \mathbf{X}_i'\beta + \epsilon_i$ and $\mathbb{E}[\epsilon_i | X_i] = 0$ (Key Concept 4.3.1 in Text, p.131)
  - **Lecture Slides:** Handout2 *(pp.9-11)*
  - **Section Notes:** Section3 *(pp.3-4)*
- Assumption 2: $\{X_i, Y_i\}_{i=1}^n$ is an i.i.d. sample (Key Concept 4.3.2)
  - **Lecture Slides:** Handout2 *(pp.12)*

  - **Section Notes:** Section3 *(pp.4)*
- Assumption 3: The matrix $\mathbb{E}\left[\mathbf{X}_i\mathbf{X}_i'\right]$ is invertible (non-singular) (Implicitly assumed in Text)
  - **Lecture Slides:** Handout2 *(pp.26-27)*
  - **Section Notes:** Section3 *(pp.5)*
- Assumption 4: The matrix $\mathbb{E}\left[\mathbf{X}_i\epsilon_i^2\mathbf{X}_i'\right]$ is non-singular (Implicitly assumed in Text)
  - **Lecture Slides:** Handout2 *(pp.41)*
  - **Section Notes:** Section3 *(pp.5)*
- Assumption 5: The random variables (X_i, Y_i) have finite fourth moments.
  - **Lecture Slides:** Handout2 *(pp.45)*
  - **Section Notes:** Section3 *(pp.5)*

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 $\boldsymbol{\beta}$.
  - **Section Notes:** Section3 *(pp.8)*
- 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.
  - **Lecture Slides:** Handout2 *(pp.18-32)*
  - **Section Notes:** Section3 *(pp.5-8)*
- 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 $\hat{\mathbf{V}} = \left(\frac{1}{n}\sum_{i=1}^n \mathbf{X}_i\mathbf{X}_i'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^n \mathbf{X}_i e_i^2\mathbf{X}_i'\right)\left(\frac{1}{n}\sum_{i=1}^n \mathbf{X}_i\mathbf{X}_i'\right)^{-1}$, where $e_i = Y_i - \mathbf{X}_i'\hat{\boldsymbol{\beta}}$, i.e., $\sqrt{n}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) \xrightarrow[n\to\infty]{d} \mathcal{N}\left(\mathbf{0}, \hat{\mathbf{V}}\right)$.
  - **Lecture Slides:** Handout2 *(pp.33-46)*
  - **Section Notes:** Section3 *(pp.5-9)*

1.5 Hypothesis Testing in Simple Regression

- Distribution of $\hat{\boldsymbol{\beta}}$
  - **Lecture Slides:** Handout3 *(pp.2)*
  - **Section Notes:** Section4 *(pp.1)*
- Definition of type I error, type II error, critical value, etc.
  - **Lecture Slides:** Handout3 *(pp.4-5)*
  - **Section Notes:** Section4 *(pp.2-3)*
- Hypothesis testing for a single parameter - t Test
  - **Lecture Slides:** Handout3 *(pp.3-6)*
  - **Section Notes:** Section4 *(pp.3)*
- Hypothesis testing for multiple parameters - Wald Test
  - **Lecture Slides:** Handout3 *(pp.7-17)*
  - **Section Notes:** Section4 *(pp.4-5)*
- Hypothesis testing for multiple parameters - F-Statistic
  - **Lecture Slides:** Handout3 *(pp.18-19)*

1.6 Confidence Intervals in Simple Regression

- Confidence interval for a single parameter
  - **Lecture Slides:** Handout3 *(pp.20)*
  - **Section Notes:** Section4 *(pp.5)*
- Confidence region for multiple parameters
  - **Lecture Slides:** Handout3 *(pp.21-24)*

1.7 Other Inference Related Concepts

- Conditional homoscedasticity
  - **Lecture Slides:** Handout3 *(pp.25)*
- Formula for the asymptotic variance under conditional homoscedasticity
  - **Lecture Slides:** Handout3 *(pp.26-28)*
- R-square
  - **Lecture Slides:** Handout3 *(pp.29-31)*
- Standard error of the regression
  - **Lecture Slides:** Handout3 *(pp.32-34)*

2 Multiple Regression

2.1 Motivation for Multiple Regression

- Interpret LR as Best Linear Predictor (BLP)
  - **Lecture Slides:** Handout4 *(pp.5-11)*
- Omitted variable bias: formula and assessment
  - **Lecture Slides:** Handout4 *(pp.12-19)*
  - **Section Notes:** Section4 *(pp.1-8)*

2.2 Multiple regression

- Population model for multiple regression and interpretation of the parameters ("holding everything else constant")
  - **Lecture Slides:** Handout4 *(pp.20-21)*
- Formula for the OLS estimator (same as in simple regression model if written in matrix form)
  - **Lecture Slides:** Handout4 *(pp.26)*
- Assumptions (same as in simple regression model, only differ in the dimensions of the matrices)
  - **Lecture Slides:** Handout4 *(pp.26)*
  - **Section Notes:** Section5 *(pp.8-9)*
- Properties of OLS estimator: consistency and asymptotic normality
  - **Lecture Slides:** Handout4 *(pp.26)*

2.3 Hypothesis Testing in Multiple Regression

- Distribution of $\hat{\boldsymbol{\beta}}$
  - **Lecture Slides:** Handout5 *(pp.2)*
- Hypothesis testing for a single parameter - t Test (same as in simple regression)
- Hypothesis testing for multiple parameters - Wald Test
  - **Lecture Slides:** Handout5 *(pp.3-8)*
  - **Section Notes:** Section6 *(pp.1-5)*

2.4 Confidence Regions in Multiple Regression

- Confidence interval for a single parameter (same as in simple regression)
- Confidence region for multiple parameters
  - **Lecture Slides:** Handout5 *(pp.9-12)*

2.5 Other Inference Related Concepts

- R-squared and adjusted R-squared, and their interpretations
  - **Lecture Slides:** Handout5 *(pp. 13-15)*
- Standard errors of the estimator in multiple regression under the additional assumption of conditional homoscedasticity
  - **Lecture Slides:** Handout5 *(pp.16-18)*

2.6 Nonlinear Regression Functions

- General non-linear regression model
  - **Lecture Slides:** Handout5 *(pp.21-22)*
  - **Section Notes:** Section6 *(pp.6)*
- Polynomials
  - **Lecture Slides:** Handout5 *(pp.23-27)*
  - **Section Notes:** Section6 *(pp.6)*
- Interaction terms: interactions between two binary variables, interactions between a continuous and a binary variable, interactions between two continuous variables
  - **Lecture Slides:** Handout6 *(pp.3-16)*
  - **Section Notes:** Section6 *(pp.7-8)*
- Dependent variable transformations: log-linear, linear-log, and log-log specifications
  - **Lecture Slides:** Handout6 *(pp.17-20)*
  - **Section Notes:** Section6 *(pp.7)*
- Application to empirical problem
  - **Lecture Slides:** Handout6 *(pp.21-32)*

2.7 Assessing Studies

- Internal validity: omitted variable bias, wrong functional form, measurement error, sample selection bias, simultaneous causality bias
  - **Lecture Slides:** Handout6 *(pp.33-52, pp.65-68)*
- External validity
  - **Lecture Slides:** Handout6 *(pp.53-64)*