Final Review

Outline

- Simple and Multiple Linear Regression (Estimation, Inference)
- Special Topics for Multiple Regression
    - Extra Sums of Squares
    - Standardized Version of the Multiple Regression Model
- Polynomial and Interaction Regression Models
- Diagnosis for Simple and Multiple Linear Regression Models
- Model Selection

- Inference in simple and multiple linear regression
  - Estimation: MLE, LSE
  - Confidence interval and prediction interval
  - Hypothesis testing: t test, ANOVA F test, General linear test
    - H_0: $\beta_k = 0$; $H_0 : \beta_1 = \beta_2 = \ldots = \beta_{p\text{-}1} = 0$; $H_0 : \beta_2 = \beta_3 = 0$;
    - $H_0 : \beta_2 = \beta_3$; $H_0 : \beta_0 = 0, \beta_1 = 2$;
    - H_0: identity (or parallel) of regression lines
    - Test for Linearity

$$H_0 : E(Y_i) = \beta_0 + \beta_1 X_i \quad \text{vs} \quad H_A : E(Y_i) = \mu_i \neq \beta_0 + \beta_1 X_i$$

- General Regression Model in Matrix Terms
  - Design matrix
  - Hat matrix

- Special Topics for Multiple Regression
    - Extra Sums of Squares
    - ANOVA table
    - Coefficients of Determination and Coefficients of Partial Determination
    - Standardized Regression Model
    - Qualitative Predictors
- Polynomial and Interaction Regression Models
    - Polynomial Regression
    - Interpretation of Regression Models with Interactions
    - Qualitative Predictors

Diagnostics

- Diagnostics for residuals (including Graphical diagnostics)
  - L.I.N.E(Linearity; Independence; Normality; Equality of variance)
- Lack of fit (test for Linearity using general linear test)
- Outlier detection:
  - Studentized (Deleted) Residuals (Identifying outlyingY)
  - Hat Matrix Leverage Values (Identifying outlying X)
  - Cook's Distance (Identifying Influential Cases)
- Multicollinearity Diagnostic
  - Variance Inflation Factor

- Model Selection
  - Six Criteria

$$R_p^2, R_{a,p}^2, C_p, AIC_p, BIC_p(SBC_p), PRESS_p$$

  - Stepwise Regression Methods

- Logistics regression
  - Odds ratio (OR)
  - Likelihood
  - Inference

Chapter 1 Simple Linear Regression

- **Concepts in Regression Models**
  - random error, residuals, fitted value, ……
- **Simple Linear Regression Model with Distribution of Error Terms Unspecified**
  - Least square estimators (LSEs)
  - Properties of LSEs
- **Normal Error Regression Model**
  - Maximum likelihood estimators (MLEs)
  - Properties of MLEs

Properties of LSEs

Under linear regression model (1.1) in which the errors have expectation zero and are uncorrelated and have equal variances σ^2.

(1) Least squares estimators (LSEs) b_0 and b_1 are linear combinations of $\{Y_i\}$

(2) (*Gauss-Markov theorem*) Least squares estimators b_0 and b_1 are BLUE (best linear unbiased estimators) of β_0 and β_1 respectively.

- Best: have minimum variance among all unbiased linear estimators

(3) MSE is an unbiased estimator of σ^2, i.e. E(MSE) $= \sigma^2$.

Properties of MLEs

In normal error regression model,

(1) MLEs of β_0 and β_1 are same with LSE estimators b_0 and b_1. They are linear combinations of $\{Y_i\}$.

(2) MLEs of β_0 and β_1 are BLUEs and normal distributed

$$\begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} \sim N\left(\begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}, \frac{\sigma^2}{SS_{XX}} \begin{pmatrix} \frac{1}{n}\sum X_i^2 & -\bar{X} \\ -\bar{X} & 1 \end{pmatrix} \right)$$

(3) MSE of σ^2 is a biased estimator with

$$\frac{n\hat{\sigma}^2}{\sigma^2} = \frac{SSE}{\sigma^2} \sim \chi^2(n-2) \quad \text{and} \quad E(\hat{\sigma}^2) = \frac{n-2}{n}\sigma^2 \to \sigma^2$$

(4) $(\hat{\beta}_0, \hat{\beta}_1, \bar{Y})$ and $\hat{\sigma}^2$ (or SSE) are independent.

Chapter 2 Inference in Simple Regression

- Inference about regression parameters in the Normal Error Regression Model
- Estimation of mean response *EY*
- Prediction Interval of New Observation
- ANVOA (Analysis of Variance) Approach to Regression Analysis
- Determination and Pearson correlation

Chapter 3, 6, 10 & 11
Diagnostics and Remedial Measures

Diagnostics (using residual plots and tests)

- L.I.N.E(Linearity; Independence; Normality; Equality of variance); Outliers; Lack of fit; Multicollinearity

- Tests involving residuals
  - Tests for constancy of variance (Brown-Forsythe test, Breusch-Pagan test, Chapter 3 & 6)
  - Tests for outliers (Chapter 10)
  - Tests for normality of error distribution

Remedial Measures

- Nonlinearity of regression function - Transformation(s) (Chapter 6) , non-parametric (Chapter 11) ~~or nonlinear regression(Chapter13)~~
- Nonconstancy of error variance - Weighted least squares (Chapter 11) and transformations (Chapter 6)
- Non-normality of error terms - Transformations (Chapter 6) or fit Generalized Linear Model(Chapter 14)
- Omission of Important Predictor Variables - Include important predictors in a multiple regression model (Chapter 6 and later on)
- Outlying observations - Robust regression (Chapter 11)
- Multicollinearity –Ridge regression(Chapter 11)

Chapter 4 & Chapter 5

Chapter 4

- Bonferroni Correction for Simultaneous Inference
- Regression Through the Origin

Chapter 5

- General Regression Model in Matrix Terms
  - Design matrix
  - Hat matrix

Chapter 6,7 &8 Multiple Regression

- Inference about regression parameters, EY and prediction
- ANVOA Approach
- Extra sums of squares
- General linear test (partial F test)
- Partial determination and partial correlation
- Standardized version of the multiple regression model
- Polynomial Regression Models
- Interaction Regression Models
    - Qualitative Predictors

Chapter 9 Model Selection and Validation

- Criteria for model selection
- Search procedures for model selection
  - Best subsets algorithm
  - Stepwise, forward,…
- Model validation

Chapter 14 Logistics regression

- Odds ratio (OR)
- Likelihood
- Inference
  - MLE estimator of coefficients
  - CI for β_k
  - Test for single or several $\beta_k = 0$.
    - H_0: $\beta_k = 0$; $\mathrm{H}_0 : \beta_2 = \beta_3 = 0$;
- Accuracy of prediction:
  - TPR(=1 − FNR) and FPR(=1 −TNR)
  - Sensitivity(=TPR) and specificity (=TNR =1 − FPR)
  - ROC curve