

STT 465

- I. Multiple Linear
- II. OLS Estimation
- III. Max. Likelihood

Multiple Linear Regression

- Linear Regression Model

$$y_i = \mu + x_{i1}\beta_1 + x_{i1}\beta_2 + \dots + x_{ip}\beta_p + \varepsilon_i$$
$$= \mu + \sum_{j=1}^p x_{ij}\beta_j + \varepsilon_i$$

- Matrix representation

Let $x'_i = (1, x_{i1}, x_{i2}, \dots, x_{ip})$ $\beta = (\mu, \beta_1, \beta_2, \dots, \beta_p)'$

Then $y_i = x'_i\beta + \varepsilon_i$

Stack equations 1 to n to get $y = X\beta + \varepsilon$

Where $y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$ $X = \begin{bmatrix} x'_1 \\ \vdots \\ x'_n \end{bmatrix}$ or $X = [x_1, \dots, x_p]$ and $\varepsilon = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$

Multiple Linear Regression

- Residual sum of squares

$$RSS = \sum_{i=1}^n \left(y_i - \mu - \sum_{j=1}^p x_{ij} \beta_j \right)^2 = (y - X\beta)' (y - X\beta)$$

- Ordinary-Least Squares (OLS)

- Take derivative of the RSS with respect to one coefficient
- Set the resulting equation equal to zero (FOC)
- Do the same for all coefficients
- This yields as many equations as unknowns, solve for the coefficients.
- We are going to stack all these FOC to get a closed-form matrix representation of the OLS solution.
- The solution will take the following form

$$[X'X] \hat{\beta} = X'y$$

or, for full-rank systems

$$\hat{\beta} = [X'X]^{-1} X'y$$

Steps for deriving OLS estimates

$$\begin{aligned}\frac{dRSS}{d\beta_j} &= -2 \sum_{i=1}^n \left(y_i - \sum_{k=1}^p x_{ik} \beta_k \right) x_{ij} \\ &= -2 \sum_{i=1}^n \left(x_{ij} y_i - \sum_{k=1}^p x_{ij} x_{ik} \beta_k \right) \\ &= -2 \left[\sum_{i=1}^n x_{ij} y_i - \sum_{i=1}^n \sum_{k=1}^p x_{ij} x_{ik} \beta_k \right] \\ &= -2 \left[\sum_{i=1}^n x_{ij} y_i - \sum_{k=1}^p \sum_{i=1}^n x_{ij} x_{ik} \beta_k \right] \\ &= -2 \left[x'_j y - \sum_{k=1}^p x'_j x_k \beta_k \right]\end{aligned}$$

$$FOC_j : -2 \left[x'_j y - \sum_{k=1}^p x'_j x_k \hat{\beta}_k \right] = 0 \Leftrightarrow \sum_{k=1}^p x'_j x_k \hat{\beta}_k = x'_j y$$

Steps for deriving OLS estimates

$$\begin{aligned}\frac{dRSS}{d\beta_j} &= -2 \sum_{i=1}^n \left(y_i - \sum_{k=1}^p x_{ik} \beta_k \right) x_{ij} \\ &= -2 \sum_{i=1}^n \left(x_{ij} y_i - \sum_{k=1}^p x_{ij} x_{ik} \beta_k \right) \\ &= -2 \left[\sum_{i=1}^n x_{ij} y_i - \sum_{i=1}^n \sum_{k=1}^p x_{ij} x_{ik} \beta_k \right] \\ &= -2 \left[\sum_{i=1}^n x_{ij} y_i - \sum_{k=1}^p \sum_{i=1}^n x_{ij} x_{ik} \beta_k \right] \\ &= -2 \left[x'_j y - \sum_{k=1}^p x'_j x_k \beta_k \right]\end{aligned}$$

$$FOC_j : -2 \left[x'_j y - \sum_{k=1}^p x'_j x_k \hat{\beta}_k \right] = 0 \Leftrightarrow \sum_{k=1}^p x'_j x_k \hat{\beta}_k = x'_j y$$

Stack all the FOCs in a system of linear equations

$$FOC_j : \sum_{k=1}^p x'_j x_k \hat{\beta}_k = x'_j y$$

$$\begin{bmatrix} x'_1 x_1 & \cdots & x'_1 x_p \\ \vdots & \ddots & \vdots \\ x'_p x_1 & \cdots & x'_p x_p \end{bmatrix} \begin{pmatrix} \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_p \end{pmatrix} = \begin{bmatrix} x'_1 y \\ \vdots \\ x'_p y \end{bmatrix}$$

$$[X'X]\hat{\beta} = X'y$$

Maximum Likelihood Estimation Under Normal Assumptions

Multiple linear regression with normal error terms

$$y_i = \sum_{j=1}^p x_{ij} \beta_j + \varepsilon_i \quad [x_{1i} = 1; \beta_1 = \mu] \quad \varepsilon_i \stackrel{iid}{\sim} N(0, \sigma_\varepsilon^2)$$

Likelihood Function

$$\begin{aligned} p(y | \beta, \sigma_\varepsilon^2) &= \prod_{i=1}^n \frac{\text{Exp} \left\{ \frac{-\left(y_i - \sum_{j=1}^p x_{ij} \beta_j\right)^2}{2\sigma_\varepsilon^2} \right\}}{\sqrt{2\pi\sigma_\varepsilon^2}} \\ &= \left(\frac{1}{2\pi\sigma_\varepsilon^2} \right)^{-n/2} \text{Exp} \left\{ \frac{-1}{2\sigma_\varepsilon^2} \sum_{i=1}^n \left(y_i - \sum_{j=1}^p x_{ij} \beta_j \right)^2 \right\} \end{aligned}$$

Maximum Likelihood Estimation Under Normal Assumptions

Likelihood Function

$$L(\beta, \sigma_\varepsilon^2 \mid y) = \left(\frac{1}{2\pi\sigma_\varepsilon^2} \right)^{-n/2} \text{Exp} \left\{ \frac{-RSS(y, \beta)}{2\sigma_\varepsilon^2} \right\}$$

Log-Likelihood Function

$$l(\beta, \sigma_\varepsilon^2 \mid y) = -\frac{n}{2} \log(2\pi\sigma_\varepsilon^2) - \frac{1}{2\sigma_\varepsilon^2} RSS(y, \beta)$$

MLE of Reg. Coefficients

$$l(\beta \mid \sigma_\varepsilon^2, y) \propto -\frac{1}{2\sigma_\varepsilon^2} RSS(y, \beta) \Rightarrow MLE = OLS$$

Sampling Distribution of OLS (& ML) Estimates

Sampling Distribution of OLS estimates

OLS estimator

$$\hat{\beta} = [X'X]^{-1} X'y$$

Expected value

$$\begin{aligned} E[\hat{\beta}] &= E[(X'X)^{-1} X'y] \\ &= (X'X)^{-1} X'E[X\beta + \varepsilon] \quad [\text{assuming } X'E(\varepsilon)=0] \\ &= (X'X)^{-1} X'X\beta \\ &= \beta \quad [\text{OLS estimates are unbiased}]. \end{aligned}$$

Sampling Distribution of OLS estimates

OLS estimator

$$\hat{\beta} = [X'X]^{-1} X'y$$

Variance

$$\text{Cov}[\hat{\beta}, \hat{\beta}'] = \sigma_{\varepsilon}^2 (X'X)^{-1} \quad [\text{derivation presented in class}]$$

Asymptotic distribution

$$\hat{\beta} \sim MVN\left[\beta, (X'X)^{-1} \sigma_{\varepsilon}^2\right]$$

Applications

- **Intercept model.**
- **Two (or more) means model.**
- **Linear regression.**
- **Linear regression with two or more groups.**

Applications

- **Intercept model.**
- **Two (or more) means model.**
- **Linear regression.**
- **Linear regression with two or more groups**