Problem Set 1

Due Sep 12, before class

1. The simplest OLS

Show that the sample mean equals

$$\bar{y} = \arg\min_{\mu} n^{-1} \sum_{i=1}^{n} (y_i - \mu)^2.$$

2. Univariate OLS without intercept

Find

$$\hat{\beta} = \arg\min_{\beta} n^{-1} \sum_{i=1}^{n} (y_i - \beta x_i)^2.$$

3. Pairwise slopes

Given $(x_i, y_i)_{i=1}^n$ with univariate x_i and y_i, show that

$$\hat{\beta} = \sum_{(i,j)} w_{ij} b_{ij},$$

where the summation is over all pairs of observations (i, j),

$$b_{ij} = \frac{y_i - y_j}{x_i - x_j}$$

is the slope determined by two points (x_i, y_i) and x_j, y_j, and

$$w_{ij} = \frac{(x_i - x_j)^2}{\sum_{(i',j')}(x_{i'} - x_{j'})^2}$$

is the weight proportional to the squared distance between x_i and x_j. In the above formulas, we define $b_{ij} = 0$ if $x_i = x_j$.

4. Invariance of OLS

Assume that $X^T X$ is non-degenerate and Γ is a $p \times p$ non-degenerate matrix. Define $\tilde{X} = X\Gamma$. From the OLS fit of Y on X, we obtain the coefficient $\hat{\beta}$, the fitted value $\hat{Y}$, the residual $\hat{\epsilon}$, and the hat matrix $\hat{H}$; From the OLS fit of Y on $\tilde{X}$, we obtain the coefficient $\tilde{\beta}$, the fitted value $\tilde{Y}$, the residual $\tilde{\epsilon}$, and the hat matrix $\tilde{H}$.

(a) Prove that

$$\hat{\beta} = \Gamma\tilde{\beta}, \quad \hat{Y} = \tilde{Y}, \quad \hat{\epsilon} = \tilde{\epsilon}, \quad \hat{H} = \tilde{H}.$$

(b) In a treatment-control experiment with m treated and n control units

$$X = \begin{pmatrix} 1_m & 1_m \\ 1_n & 0_n \end{pmatrix},$$

show that $H = \mathrm{diag}\{m^{-1}1_m 1_m^T, n^{-1}1_n 1_n^T\}$.

5. Full sample and subsample OLS

Partition the full sample into K subsamples:

$$X = \begin{pmatrix} X_{(1)} \\ \vdots \\ X_{(K)} \end{pmatrix}, \quad Y = \begin{pmatrix} Y_{(1)} \\ \vdots \\ Y_{(K)}, \end{pmatrix}$$

where the k-th sample consists of $(X_{(k)}, Y_{(k)})$ with $X_{(k)} \in \mathbb{R}^{n_k \times p}$ and $Y_{(k)} \in \mathbb{R}^{n_k}$ being the covariate matrix and outcome vector. Let $\hat{\beta}$ be the OLS coefficient based on the full sample, and $\hat{\beta}_{(k)}$ be OLS coefficient based on the k-th sample. Show that

$$\hat{\beta} \;=\; \sum_{k=1}^{K} W_{(k)} \hat{\beta}_{(k)},$$

where the weight matrix equals

$$W_{(k)} \;=\; \left(\sum_{k'=1}^{K} X_{(k')}^T X_{(k')} \right)^{-1} X_{(k)}^T X_{(k)}.$$