EML Assignment 1 Problem 3

Muhammed Saeed and Ali Salaheldin Ali Ahmed

November 2022

Question 1

The Gauss-Markov theorem states that the least square-fit has the least variance among all linear unbiased estimators for the population regression line.

$$y_i = \sum_{j=1}^{K} \beta_j x_{ij} + \epsilon_i \tag{1}$$

Regression analysis is like any other inferential methodology, in which the goal is to draw a random sample from a population and use it to estimate the properties of that population. The theorem is important because it tells us that the least-squares estimator parameters $\hat{\beta}_j$ are most likely to be the closest to the actual parameters β_j of the population regression line, since it is unbiased (i.e. $\mathbb{E}[\hat{\beta}_j] = \beta_j$) and has the least variance.

Question 2

- The expected value of the errors is zero, since otherwise the estimator would be biased such that the expected value of the model parameters $\mathbb{E}[\hat{\beta}|X]$ would not be equal to population parameter, because the samples do not necessarily represent the population.

$$\mathbb{E}(\epsilon_i) = 0$$

- The error terms must be independent, because any correlation between the errors with the inputs or with other error terms would introduce bias.

$$\mathrm{Cov}(\epsilon_i, x_j) = 0, \forall i, j$$
$$\mathrm{Cov}(\epsilon_i, \epsilon_j) = 0, \forall i \neq j$$

- The variance for all error terms must be finite and equal. This to ensure that the model is able to estimate the population from a finite size sample.

$$\mathrm{Var}(\epsilon_i) = \sigma^2 < \infty, \forall i$$

Question 3

The least-squares estimator is also the best unbiased linear estimator in terms of test error. This is because, while all unbiased linear estimators have the same bias level, the least-squares estimator has the least variance.