Statistics

Harry Han

March 31, 2023

1 Linear Regression

1.1 Least Square Estimate

1.1.1 Single Variabel Linear Regression

Theorema 1.1.1 (Least Square Estimators)**.** *Assuming iid random variables x_i with the regression model*

$$Y_i = \alpha + \beta x_i + \epsilon_i \tag{1}$$

With the assumption that each ϵ_i is normally distributed, $\mathbb{E}[\epsilon_i] = 0$ (Linearity of Expectation), and $Var(\epsilon_i) = \sigma^2$ (homoscedasticity)

The least square estimator are

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{x}; \hat{\beta} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(x_i - \bar{x}^2)} \tag{2}$$

Both estimator are consistent and unbiased. Moreover:

$$Var(\hat{\beta}) = \frac{\sigma^2}{(n-1)s_x^2}; Var(\hat{\alpha}) = \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{(n-1)s_x^2}\right) \tag{3}$$

Where $\sigma^2 = Var(\epsilon_i) = Var(Y_i)$ and $s_x^2 = \frac{\sum(x_i - x)^2}{n-1}$ is the sample variance of the explanatory varaible.

Observatio 1.1 (Assumption of Linear Regression model)*.* There are four assumption for the linear regression model:

1. Linearity of Expectation: $\mathbb{E}[Y_i] = \alpha + \beta x_i$
2. Homoscedasticity (same variance): $\text{Var}(\epsilon_i) = \sigma^2$

3. Independence: ϵ_i and x_i are independent

4. Normality: ϵ_i is normally distributed

Definitio 1.1 (Residue Sum of Square)**.** The residue sun of square, SSE, is defined as:

$$SSE = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{4}$$

Its unbiased estimator is $S_E^2 = \frac{1}{2}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$

The regression equation for the response given the explanatory varaiable x_0 is

$$\mathbb{E}[Y_0] = \alpha + \beta x_0 \tag{5}$$