GAUSS-MARKOV THEOREM

We have seen that we can derive the estimate $\hat{\beta}$ both as a maximization of the likelihood under the Gaussian linear model assumption (ML estimate) and as a minimization of the sum of squares (OLS estimate), without the need to specify a distribution (and only using conditions on the first two moments).

We consider now the second framework $\rightarrow$ remove the distributive assumption.

Assume that: 1) $\underline{Y} = X\underline{\beta} + \underline{\varepsilon}$ linearity

2) $\mathbb{E}[\underline{\varepsilon}] = \underline{0}$ and $\mathrm{var}(\underline{\varepsilon}) = \sigma^2 I_n$ (homoscedasticity and incorrelation)

3) X non-stochastic with full rank $(\mathrm{rank}(X) = p)$

The OLS estimator is $\underline{\hat{B}} = (X^T X)^{-1} X^T \underline{Y}$ (linear transformation of $\underline{Y}$)

Even without the specification of a distribution for $\underline{Y}$, we can still derive the first two moments of $\underline{\hat{B}}$.

We have already computed them: $\underset{\text{unbiased}}{\mathbb{E}[\underline{\hat{B}}] = \beta}$, $\mathrm{var}(\underline{\hat{B}}) = (X^T X)^{-1} \sigma^2$.

GAUSS-MARKOV THM.

Consider the framework defined by assumptions (1) (2) (3).

Then the OLS estimator $\underline{\hat{B}}$ is B.L.U.E. (i.e. the Best Linear Unbiased Estimator)

"best" = "minimum variance"

So the theorem states that, in the class of linear and unbiased estimators of $\underline{\beta}$, $\underline{\hat{B}}$ has the minimum variance.

(notice however that it doesn't mean that $\underline{\hat{B}}$ is "the best estimator overall", it is the best only if we restrict to the class of linear and unbiased).

Assume that $\underline{\tilde{B}}$ is another linear unbiased estimator.

(i.e. $\underline{\tilde{B}} = A \cdot \underline{Y}$ and $\mathbb{E}[\underline{\tilde{B}}] = \beta$)

The thm states that

$$\mathrm{var}(\underline{\tilde{B}}) \geq \mathrm{var}(\underline{\hat{B}})$$