# Mean and Variance of Least Squares Estimators (Full Column Rank Case)

Arunav Bhowmick (BS2025)

August 2022

We are dealing with the Gauss-Markov set up which can be briefly described as follows-

$$\vec{y}_{n\times 1} = \mathbf{X}_{n\times p}\vec{\beta}_{p\times 1} + \vec{\epsilon}_{n\times 1}, \tag{1}$$

where, $\vec{\epsilon}_{n\times 1} \sim (\vec{0}_{n\times 1}, \sigma^2\mathbf{I}_{n\times n})$ is the Gauss-Markov assumption. Basically, the model assumes that the random error vector $\vec{\epsilon}$ has mean $\vec{0}$ and the covariance matrix is $\sigma^2\mathbf{I}$. This captures two things in one notation; the variances are the same (all the diagonal entries are $\sigma^2$) and all the covariances are zero (all the off-diagonal entries are 0) and thus the errors are uncorrelated. Note that the assumptions are only on the first two moments of $\vec{\epsilon}$.

Now, we shall see that if the Gauss-Markov assumption is met, then the least squares estimator has certain useful properties for any linear model. We will also make one more simplifying assumption which we would relax later. We assume that the design matrix $\mathbf{X}$ is full column rank i.e., $\mathbf{X}'\mathbf{X}$ is an invertible matrix. In this case, the least squares estimator for $\vec{\beta}$ is unique and is given by-

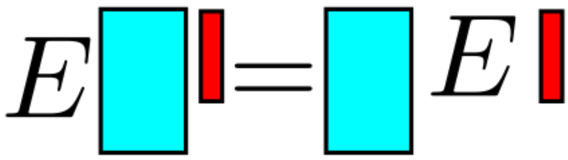
$$\hat{\vec{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\vec{y} \tag{2}$$

We will explore various interesting properties of this estimator. Observe that this estimator is a linear function of $\vec{y}$. It is obtained by premutliplying a fixed known matrix to $\vec{y}$. We would like to find the expectation and variance (by variance we mean the covariance matrix) of this estimator.

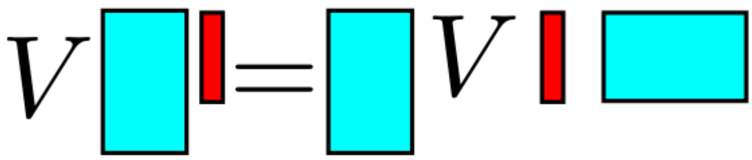
For this purpose, it might be helpful to recall how we compute the expectation and variance of a random vector obtained by premultiplying a fixed matrix of appropriate order to a random vector with known expectation and variance. Let $\mathbf{A}$ be an order $n \times p$ matrix and $\vec{\mu}$ be an order $p \times 1$ random vector. Then-

$$E(\mathbf{A}\vec{\mu}) = \mathbf{A}E(\vec{\mu}) \tag{3}$$
$$V(\mathbf{A}\vec{\mu}) = \mathbf{A}V(\vec{\mu})\mathbf{A}' \tag{4}$$

We can also represent this pictorially. Let $\vec{\mu}$ be denoted by a red rectangle and $\mathbf{A}$ be denoted by a cyan rectangle. Then-




and,



Now, we apply these results on our estimator to find out its expectation and variance. But before that, we do some auxiliary calculations which would come in handy for this task.

$$\begin{aligned} E(\vec{y}) &= E(\mathbf{X}\vec{\beta} + \vec{\epsilon}) \\ &= E(\mathbf{X}\vec{\beta}) + E(\vec{\epsilon}) \\ &= \mathbf{X}\vec{\beta} \end{aligned} \tag{5}$$

Observe that we wrote $E(\mathbf{X}\vec{\beta}) = \mathbf{X}\vec{\beta}$. Although $\mathbf{X}\vec{\beta}$ is unknwon, it is a fixed vector i.e., it is not random. Hence, its expectation will equal itself. Similarly,

$$\begin{aligned} V(\vec{y}) &= V(\mathbf{X}\vec{\beta} + \vec{\epsilon}) \\ &= V(\vec{\epsilon}) \\ &= \sigma^2\mathbf{I} \end{aligned} \tag{6}$$

Finally, we compute the expectation and variance of $\hat{\vec{\beta}}$.

$$\begin{aligned} E(\hat{\vec{\beta}}) &= E\left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\vec{y}\right] \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\left[E(\vec{y})\right] \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\vec{\beta}) \\ &= (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{X})\vec{\beta} \\ &= \vec{\beta} \end{aligned} \tag{7}$$

$$
\begin{aligned}
V(\hat{\vec{\beta}}) &= V\left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\vec{y}\right] \\
&= \left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right] V(\vec{y}) \left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right]' \\
&= \left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right] \sigma^2\mathbf{I} \left[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\right] \\
&= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1} \\
&= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}
\end{aligned} \tag{8}
$$

One interesting fact about the variance stems from our assumption that $\mathbf{X}'\mathbf{X}$ is non-singular. If it so happens that it is very nearly singular, or in other words, its determinant is very close to zero, then the variance will blow up because it has the determinant term in the denominator. This is expected since if it is exactly singular, then we would not get a unique $\hat{\vec{\beta}}$. So, when it is very nearly singular, we can expect $\hat{\vec{\beta}}$ to approach non-uniqueness i.e., it is highly variable.