5 Least Squares Estimation

Recall the linear model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$:

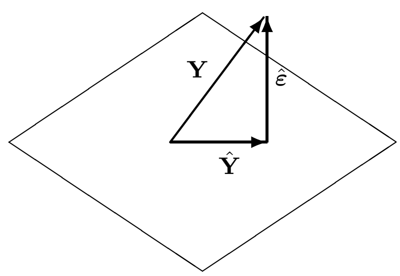
$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} x_{10} & x_{11} & \cdots & x_{1,p-1} \\ x_{20} & x_{21} & \cdots & x_{2,p-1} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n0} & x_{n1} & \cdots & x_{n,p-1} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

5.1 Definition: An estimate $\hat{\boldsymbol{\beta}}$ is a least-squares estimate of $\boldsymbol{\beta}$ if it minimizes the length $||\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}||$ over all $\boldsymbol{\beta}$.

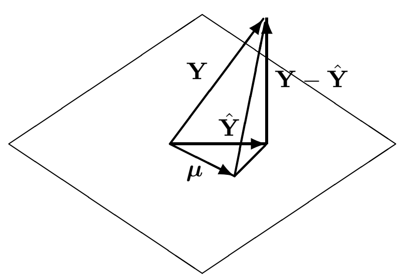
5.2 Note: Let $\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_{p-1}$ be the columns of $\mathbf{X}$. Then $\mathbf{X}\boldsymbol{\beta} = \beta_0\mathbf{x}_0 + \beta_1\mathbf{x}_1 + \ldots \beta_{p-1}\mathbf{x}_{p-1} \in \mathcal{R}(\mathbf{X})$, the range (column space) of $\mathbf{X}$. Hence a least-squares estimate can be found by minimizing $||\mathbf{Y} - \boldsymbol{\mu}||$ over $\boldsymbol{\mu} \in \mathcal{R}(\mathbf{X})$.

5.3 Theorem: $\mathbf{Y}$ can be uniquely decomposed as $\mathbf{Y} = \hat{\mathbf{Y}} + \hat{\varepsilon}$ where $\hat{\mathbf{Y}} \in \mathcal{R}(\mathbf{X})$, $\hat{\boldsymbol{\epsilon}} \in [\mathcal{R}(\mathbf{X})]^{\perp}$, and $[\mathcal{R}(\mathbf{X})]^{\perp}$ is the orthogonal complement of $\mathcal{R}(\mathbf{X}) = \{\mathbf{a} : \mathbf{X}'\mathbf{a} = \mathbf{0}\}$.

5.4 Definition: $\hat{\mathbf{Y}}$ (sometimes written as $\hat{\boldsymbol{\mu}}$) is the orthogonal projection of $\mathbf{Y}$ onto $\mathcal{R}(\mathbf{X})$. It is also called the fitted vector or vector of fitted values. The residual vector is $\hat{\boldsymbol{\varepsilon}} = \mathbf{Y} - \hat{\mathbf{Y}} = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}$.



5.5 Theorem: The orthogonal projection solves the least-squares minimization problem.



5.6 Theorem: A least squares estimate is a solution to the normal equations: $\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{X}'\mathbf{Y}$.

5.7 Definition: The residual sum of squares is defined by

$$RSS = \hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}} = (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = \mathbf{Y}'\mathbf{Y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}}.$$

5.8 Theorem: If $\text{rank}(\mathbf{X}^{n\times p}) = p$, then $\text{rank}(\mathbf{X}'\mathbf{X}) = p$, so $(\mathbf{X}'\mathbf{X})^{-1}$ exists. In this case the normal equations have the unique solution

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}.$$

The orthogonal projection is

$$\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \mathbf{P}\mathbf{Y},$$

where

$$\mathbf{P} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'.$$

5.9 Theorem: Let $\text{rank}(\mathbf{X}^{n\times p}) = p$, and $\mathbf{P} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. Then

(a) $\mathbf{P}$ and $\mathbf{I} - \mathbf{P}$ are projection matrices.

(b) $\text{rank}(\mathbf{I} - \mathbf{P}) = \text{tr}(\mathbf{I} - \mathbf{P}) = n - p$.

(c) $\mathbf{P}\mathbf{X} = \mathbf{X}$.

5.10 Note: $\mathbf{P}$ is projection onto $\mathcal{R}(\mathbf{X})$. $\mathbf{I} - \mathbf{P}$ is projection onto $[\mathcal{R}(\mathbf{X})]^{\perp}$. The residual vector becomes

$$\hat{\boldsymbol{\varepsilon}} = \mathbf{Y} - \hat{\mathbf{Y}} = (\mathbf{I} - \mathbf{P})\mathbf{Y},$$

and the residual sum of squares

$$RSS = \hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}} = \mathbf{Y}'(\mathbf{I} - \mathbf{P})\mathbf{Y}.$$

5.11 Definition: For $\mathbf{A}_{m\times n}$, a generalized inverse of $\mathbf{A}$ is an $n \times m$ matrix $\mathbf{A}^-$ satisfying $\mathbf{A}\mathbf{A}^-\mathbf{A} = \mathbf{A}$.

5.12 Theorem: In general, the projection onto $\mathcal{R}(\mathbf{X})$ is $\mathbf{P}\mathbf{Y}$, where $\mathbf{P} = \mathbf{X}(\mathbf{X}'\mathbf{X})^-\mathbf{X}'$.