8 Orthogonal Structure in the Design Matrix

Partition the linear model as

$$\boldsymbol{\mu} = (\mathbf{X}_0, \mathbf{X}_1, \cdots, \mathbf{X}_k) \begin{pmatrix} \boldsymbol{\beta}_0 \\ \boldsymbol{\beta}_1 \\ \vdots \\ \boldsymbol{\beta}_k \end{pmatrix},$$

where $\mathbf{X}_j$ is $n \times p_j$, $\boldsymbol{\beta}_j$ is $p_j \times 1$, and $\sum_j p_j = p$. Suppose that the columns of $\mathbf{X}_i$ are orthogonal to those of $\mathbf{X}_j$, i.e., $\mathbf{X}_i'\mathbf{X}_j = \mathbf{0},$ for all i, j. Then $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ has the form

$$\begin{pmatrix} \hat{\boldsymbol{\beta}}_0 \\ \hat{\boldsymbol{\beta}}_1 \\ \vdots \\ \hat{\boldsymbol{\beta}}_k \end{pmatrix} = \begin{pmatrix} (\mathbf{X}_0'\mathbf{X}_0)^{-1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & (\mathbf{X}_1'\mathbf{X}_1)^{-1} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & (\mathbf{X}_k'\mathbf{X}_k)^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{X}_0'\mathbf{Y} \\ \mathbf{X}_1'\mathbf{Y} \\ \vdots \\ \mathbf{X}_k'\mathbf{Y} \end{pmatrix} = \begin{pmatrix} (\mathbf{X}_0'\mathbf{X}_0)^{-1}\mathbf{X}_0'\mathbf{Y} \\ (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{Y} \\ \vdots \\ (\mathbf{X}_k'\mathbf{X}_k)^{-1}\mathbf{X}_k'\mathbf{Y} \end{pmatrix}.$$

Therefore, the least-squares estimate of $\boldsymbol{\beta}_i$ does not depend on whether any of the other terms are in the model. Also,

$$RSS = \mathbf{Y}'\mathbf{Y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{Y} = \mathbf{Y}'\mathbf{Y} - \sum_{i=0}^{k} \hat{\boldsymbol{\beta}}_i'\mathbf{X}_i'\mathbf{Y},$$

i.e. if $\boldsymbol{\beta}_i$ is set equal to 0, RSS increases by $\hat{\boldsymbol{\beta}}_i'\mathbf{X}_i'\mathbf{Y}$.

8.1 Example: (Simple linear regression).

- Model 1A: $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$.
  The least-squares estimate of β_1 is

$$\hat{\beta}_1 = \sum_{i=1}^{n}(x_i - \bar{x})Y_i / \sum_{i=1}^{n}(x_i - \bar{x})^2.$$

- Model 1B: $Y_i = \beta_1 x_i + \varepsilon_i$.
  The least-squares estimate of β_1 is

$$\hat{\beta}_1 = \sum_{i=1}^{n} x_i Y_i / \sum_{i=1}^{n} x_i^2.$$

The slope estimates in Models 1A and 1B are equal only when $\bar{x} = 0$, i.e., when $x = (x_1, \ldots, x_n)'$ is orthogonal to the intercept $\mathbf{1} = (1, \ldots, 1)'$. Note that in Model 1A, $\hat{\beta}_1$ and $\hat{\beta}_0$ are uncorrelated when $\bar{x} = 0$.

- Model 2A: $Y_i = \beta_0 + \beta_1\bar{x} + \beta_1(x_i - \bar{x}) + \varepsilon_i = \beta_0^* + \beta_1(x_i - \bar{x}) + \varepsilon_i$.
  Now
  $$\mathbf{X} = \begin{pmatrix} 1 & x_1 - \bar{x} \\ 1 & x_2 - \bar{x} \\ \vdots & \vdots \\ 1 & x_n - \bar{x} \end{pmatrix} = (\mathbf{x}_0, \mathbf{x}_1)$$
  has orthogonal columns, and hence
  $$\hat{\beta}_0^* = \frac{\mathbf{x}_0'\mathbf{Y}}{\mathbf{x}_0'\mathbf{x}_0} = \bar{Y}, \quad \hat{\beta}_1 = \frac{\mathbf{x}_1'\mathbf{Y}}{\mathbf{x}_1'\mathbf{x}_1} = \sum_{i=1}^{n}(x_i - \bar{x})Y_i / \sum_{i=1}^{n}(x_i - \bar{x})^2.$$
- Model 2B: $Y_i = \beta_1(x_i - \bar{x}) + \varepsilon_i$.
  $\hat{\beta}_1$ is the same as in Model 2A because of orthogonality.

8.2 Theorem: If $\mathbf{A}$ and $\mathbf{D}$ are symmetric and all inverses exist,
$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}' & \mathbf{D} \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}\mathbf{E}^{-1}\mathbf{B}'\mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{B}\mathbf{E}^{-1} \\ -\mathbf{E}^{-1}\mathbf{B}'\mathbf{A}^{-1} & \mathbf{E}^{-1} \end{pmatrix},$$
where $\mathbf{E} = \mathbf{D} - \mathbf{B}'\mathbf{A}^{-1}\mathbf{B}$.

8.3 Theorem: Assume the linear model $\mathbf{Y} \sim N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$, where the columns of $\mathbf{X}$ are linearly independent and satisfy $\mathbf{x}_i'\mathbf{x}_i \le c_i^2$ for fixed constants c_i. Then
$$\mathrm{var}(\hat{\beta}_i) \ge \sigma^2/c_i^2,$$
and the minimum is attained when $\mathbf{x}_i'\mathbf{x}_i = c_i^2$ and the columns of $\mathbf{X}$ are orthogonal, i.e., $\mathbf{x}_i'\mathbf{x}_j = 0$, for $j \ne i$.

8.4 Example: (2^k factorial design). Suppose that k factors are to be studied to determine their effect on the output of a manufacturing process. Each factor is to be varied within a given plausible range of values and the variables have been scaled so that the range is -1 to $+1$. Then the theorem implies that the optimal design has orthogonal columns and all variables set to $+1$ or -1. If $n = 2^k$ such a design is called a 2^k factorial design.