Algunas propiedades de la matriz $\mathbf{P}$:

Lema:

i) $\mathbf{P}$ y $\mathbf{I}-\mathbf{P}$ son simétricas e idempotentes

ii) $\text{rg}(\mathbf{P}) = \text{tr}(\mathbf{P}) = p$ y $\text{rg}(\mathbf{I}-\mathbf{P}) = \text{tr}(\mathbf{I}-\mathbf{P}) = n-p$

iii) $(\mathbf{I}-\mathbf{P})\mathbf{X} = \mathbf{0}$

Proposición: Dados $1 \leq i, j \leq n$ tenemos que

i) $0 \leq p_{ii} \leq 1$

ii) $-\frac{1}{2} \leq p_{ij} \leq \frac{1}{2}$ si $i \neq j$

Como ya vimos $Var(\widehat{y}_i) = \sigma^2 p_{ii}$, una consecuencia inmediata es que

$$Var(\widehat{y}_i) \leq Var(y_i) = \sigma^2 \,.$$

Una propiedad interesante es que $\mathbf{P}$ es invariante por transformaciones lineales no singulares de la forma $\mathbf{X} \to \mathbf{XA}$, donde $\mathbf{A} \in \Re^{p\times p}$ y $\mathrm{rg}(\mathbf{A}) = p$. Este tipo de transformaciones es útil, por ejemplo, si queremos realizar un cambio de unidades en las covariables.

Respecto a las propiedades de invariancia, podemos ver que si

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

para $\mathbf{A} \in \Re^{p\times p}$ no singular, $\lambda \in \Re$ y $\gamma \in \Re^{p}$, entonces

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}(\mathbf{XA}, \mathbf{Y}) &= \mathbf{A}^{-1}\widehat{\boldsymbol{\beta}} && \text{Invariancia por transformciones afines} \\
\widehat{\boldsymbol{\beta}}(\mathbf{X}, \lambda\mathbf{Y}) &= \lambda\widehat{\boldsymbol{\beta}} && \text{Invariancia por cambios de escala} \\
\widehat{\boldsymbol{\beta}}(\mathbf{X}, \mathbf{Y} + \mathbf{X}\gamma) &= \widehat{\boldsymbol{\beta}} + \gamma && \text{Invariancia por cambios de regresión}
\end{aligned}$$

Estimación de σ^2

Las varianzas de los estimadores dependen del diseño y σ^2, que es desconocida. Dado que $\sigma^2 = E(\epsilon^2)$, parece natural estimarla mediante el promedio de los cuadrados de los residuos. El vector de residuos es

$$\begin{aligned} \mathbf{r} &= \mathbf{Y} - \widehat{\mathbf{Y}} \\ &= \mathbf{Y} - \mathbf{PY}\,, \end{aligned}$$

Bajo el modelo Ω, tenemos que

$$\boxed{s^2 = \frac{\|\mathbf{Y} - \widehat{\mathbf{Y}}\|^2}{n-p} = \frac{\|\mathbf{Y} - \mathbf{PY}\|^2}{n-p}}$$

es un estimador insesgado de σ^2.

Lema Auxiliar: Sea $\mathbf{x}$ un vector aleatorio n–dimensional y sea $\mathbf{A} \in \Re^{n\times n}$ una matriz simétrica. Si $E(\mathbf{x}) = \boldsymbol{\mu}$ y su matriz de covarianza es $\Sigma_{\mathbf{x}}$ entonces

$$E(\mathbf{x}'\mathbf{A}\mathbf{x}) = tr(\mathbf{A}\Sigma) + \boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu}$$

Respecto del diseño

• Covariables aleatorias

Si las covariables son aleatorias suponemos que tenemos los vectores $(\mathbf{x}_i, y_i)$ i.i.d. que satisfacen el modelo

$$y_i = \mathbf{x}_i\boldsymbol{\beta} + \epsilon_i$$

donde los ϵ_i son i.i.d., con $E(\epsilon_i) = 0$ y $Var(\epsilon_i) = \sigma^2$ e independientes de $\mathbf{x}_i \sim F$.

El análogo de suponer que $\mathbf{X}$ tiene rango completo es asumir que la distribución de $\mathbf{x}$ no está concentrada en ningún hiperplano, es decir

$$P(\mathbf{a}'\mathbf{x} = 0) < 1 \quad \forall \mathbf{a} \neq \mathbf{0}$$

Esta condición se cumple, por ejemplo, si $\mathbf{x}$ tiene densidad.

En este caso, $\widehat{\boldsymbol{\beta}}$ está bien definido y las fórmulas que vimos para esperanza y varianza de $\widehat{\boldsymbol{\beta}}$ son válidas condicionalmente:

$$\boxed{E(\widehat{\boldsymbol{\beta}}|\mathbf{X} = \mathbf{x}) = \boldsymbol{\beta} \qquad \Sigma_{\widehat{\boldsymbol{\beta}}|\mathbf{X}=\mathbf{x}} = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}}$$

Se puede ver que si $\mathbf{V}_{\mathbf{x}} = E(\mathbf{x}\mathbf{x}')$ existe, entonces para n grande la distribución aproximada de $\widehat{\boldsymbol{\beta}}$ será

$$N_p\left(\boldsymbol{\beta}, \frac{\sigma^2 \mathbf{V}_{\mathbf{x}}^{-1}}{n}\right)$$

Cuando el modelo tiene intercept, podemos escribirlo como:

$$y_i = \beta_0 + \mathbf{x}_i' \boldsymbol{\beta}_1 + \epsilon_i$$

donde β_0 es la intercept y $\boldsymbol{\beta}_1$ es el vector de pendientes. En este caso resulta

$$\sigma^2 \mathbf{V}_{\mathbf{x}}^{-1} = \sigma^2 \begin{pmatrix} 1 + \boldsymbol{\mu}_{\mathbf{x}}' \Sigma_{\mathbf{x}}^{-1} \boldsymbol{\mu}_{\mathbf{x}} & -\boldsymbol{\mu}_{\mathbf{x}}' \Sigma_{\mathbf{x}}^{-1} \\ -\Sigma_{\mathbf{x}}^{-1} \boldsymbol{\mu}_{\mathbf{x}} & \Sigma_{\mathbf{x}}^{-1} \end{pmatrix}$$

con $\boldsymbol{\mu}_{\mathbf{x}} = E(\mathbf{x})$ y $\Sigma_{\mathbf{x}}$ matriz de covarianza de $\mathbf{x}$.

• Estructura Ortogonal en la matriz de Diseño

Supongamos que podemos dividir a la matriz $\mathbf{X}$ en k conjuntos de columnas ortogonales: $\mathbf{X}_1, \ldots, \mathbf{X}_k$, de manera que

$$\mathbf{X} = [\mathbf{X}_1 \ldots \mathbf{X}_k]$$

La correspondiente división en los parámetros daría

$$\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_k)'$$

Luego podemos escribir:

$$E(\mathbf{Y}) = \mathbf{X}_1\boldsymbol{\beta}_1 + \ldots + \mathbf{X}_k\boldsymbol{\beta}_k$$

Como las columnas de $\mathbf{X}_i$ son ortogonales a las de $\mathbf{X}_j$ si $i \neq j$, tenemos que $\mathbf{X}_i'\mathbf{X}_j = 0$, luego

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \begin{pmatrix} \mathbf{X}_1'\mathbf{X}_1 & 0 & \dots & 0 \\ 0 & \mathbf{X}_2'\mathbf{X}_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \mathbf{X}_k'\mathbf{X}_k \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{X}_1'\mathbf{Y} \\ \mathbf{X}_2'\mathbf{Y} \\ \dots \\ \mathbf{X}_k'\mathbf{Y} \end{pmatrix}$$

entonces

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \begin{pmatrix} (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{Y} \\ (\mathbf{X}_2'\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathbf{Y} \\ \dots \\ (\mathbf{X}_k'\mathbf{X}_k)^{-1}\mathbf{X}_k'\mathbf{Y} \end{pmatrix} = \begin{pmatrix} \widehat{\boldsymbol{\beta}}_1 \\ \widehat{\boldsymbol{\beta}}_2 \\ \dots \\ \widehat{\boldsymbol{\beta}}_k \end{pmatrix}$$

en consecuencia el estimador de $\boldsymbol{\beta}_i$ no cambiará si alguno de los otros $\boldsymbol{\beta}_j$ se iguala a 0, es decir si se remueve del modelo.

¿Cómo resulta la suma de cuadrados?

$$\mathbf{Y}'\mathbf{Y} - \widehat{\mathbf{Y}}'\widehat{\mathbf{Y}} = \mathbf{Y}'\mathbf{Y} - \widehat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{Y} = \mathbf{Y}'\mathbf{Y} - \sum_{j=1}^{k} \widehat{\boldsymbol{\beta}}_j'\mathbf{X}_j'\mathbf{Y}$$

Por lo tanto si en el modelo ponemos algún $\boldsymbol{\beta}_i = 0$, el único cambio en la suma de cuadrados es que el término de $\widehat{\boldsymbol{\beta}}_i'\mathbf{X}_i'\mathbf{Y}$ no aparece:

$$\mathbf{Y}'\mathbf{Y} - \sum_{\substack{j=1 \\ j \neq i}}^{k} \widehat{\boldsymbol{\beta}}_j'\mathbf{X}_j'\mathbf{Y}$$

En el caso más sencillo, cada $\mathbf{X}_i$ consta de una única columna y resulta:

$$\widehat{\boldsymbol{\beta}}_i = \frac{\mathbf{X}_i'\mathbf{Y}}{\mathbf{X}_i'\mathbf{X}_i}$$

y la suma de cuadrados queda

$$\mathbf{Y}'\mathbf{Y} - \sum_{j=1}^{k} \widehat{\boldsymbol{\beta}}_j'\mathbf{X}_j'\mathbf{Y} = \mathbf{Y}'\mathbf{Y} - \sum_{j=1}^{k} \widehat{\boldsymbol{\beta}}_j^2\mathbf{X}_j'\mathbf{X}_j$$

Teorema de Gauss–Markov

En muchas aplicaciones estamos más interesado en estimar funciones lineales de $\boldsymbol{\beta}$ que en estimar $\boldsymbol{\beta}$ en sí mismo.

Estas funciones incluyen el valor esperado de y en una futura observación $\mathbf{x}_o$, por ejemplo.

Si bien puede haber muchos estimadores de una función lineal $\mathbf{c}'\boldsymbol{\beta}$ o $\mathbf{C}\boldsymbol{\beta}$, estudiaremos los estimadores lineales, es decir funciones lineales de las observaciones $y_1, \ldots, y_n$.

Primero veremos cuando una función paramétrica es estimable.

Definición: Una función paramétrica ψ se dice que es una función lineal de los paramétros $\{\beta_1, \ldots, \beta_p\}$ si existen $\{c_1, \ldots, c_p\}$ constantes conocidas tal que

$$\psi = \mathbf{c}'\boldsymbol{\beta} = \sum_{j=1}^{p} c_j \beta_j$$

donde $\mathbf{c} = (c_1, \ldots, c_p)'$.

Definición: Decimos que una función paramétrica $\psi = \mathbf{c}'\boldsymbol{\beta}$ es estimable si tiene un estimador lineal (en $\mathbf{Y}$) insesgado, es decir si existe $\mathbf{a} \in \Re^n$ tal que

$$E(\mathbf{a}'\mathbf{Y}) = \psi = \mathbf{c}'\boldsymbol{\beta} \quad \forall \boldsymbol{\beta} \in \Re^p$$

¿Hay funciones que no son estimables?

Veamos un ejemplo de una función paramétrica no estimable.

Supongamos que queremos comparar la respuesta media de dos tratamientos y un control y que para ello observamos

$$\begin{array}{lll} \text{T1:} & y_{11}, y_{12}, \ldots, y_{1k} & y_{1j} \sim N(\beta_1, \sigma^2) \\ \text{T2:} & y_{21}, y_{22}, \ldots, y_{2k} & y_{2j} \sim N(\beta_2, \sigma^2) \\ \text{Co:} & y_{31}, y_{32}, \ldots, y_{3k} & y_{3j} \sim N(\beta_3, \sigma^2) \end{array}$$

Suponemos igual cantidad de observaciones por tratamiento para simplificar la notación.

Podemos escribir esto como

$$y_{ij} = \beta_i + \epsilon_{ij}$$

Podríamos escribir esto como un modelo lineal:

$$\mathbf{Y} = \begin{pmatrix} y_{11} \\ y_{12} \\ \cdots \\ y_{1k} \\ y_{21} \\ y_{22} \\ \cdots \\ y_{2k} \\ y_{31} \\ y_{32} \\ \cdots \\ y_{3k} \end{pmatrix} ; \mathbf{X} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ . & . & . \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ . & . & . \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ . & . & . \\ 0 & 0 & 1 \end{pmatrix} ; \boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix}$$

Por ejemplo, T1, T2 y el control podrían ser distintas dosis de una droga de manera que T1 es menor que la dosis del control y T2 mayor que la dosis control. Tendría sentido preguntarse si

$$\beta_3 = \frac{\beta_1 + \beta_2}{2}$$

lo que implicaría cierta linealidad en el efecto medio. En ese caso nos interesaría saber si

$$\left(-\tfrac{1}{2}, -\tfrac{1}{2}, 1 \right) \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} = 0$$

Otra manera de escribir el modeo sería

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij}$$

donde:

μ es el efecto general

α_i es el efecto del tratamiento i

En ese caso tendríamos

$$\mathbf{Y} = \begin{pmatrix} y_{11} \\ y_{12} \\ \cdots \\ y_{1k} \\ y_{21} \\ y_{22} \\ \cdots \\ y_{2k} \\ y_{31} \\ y_{32} \\ \cdots \\ y_{3k} \end{pmatrix} ; \mathbf{X} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ . & . & . & . \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ . & . & . & . \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ . & . & . & . \\ 1 & 0 & 0 & 1 \end{pmatrix} ; \boldsymbol{\beta} = \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix}$$

¿Son todas las funciones estimables en este modelo?

Consideremos

$$\alpha_1 = \begin{pmatrix} 0, 1, 0, 0 \end{pmatrix} \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix}$$

Veremos que α_1 no es estimable.

Veamos el siguiente resultado que caracteriza las funciones paramétricas estimables.

Teorema: La función paramétrica $\psi = \mathbf{c}'\boldsymbol{\beta}$ es estimable si y sólo si $\mathbf{c}$ es una combinación lineal de las filas de $\mathbf{X}$, o sea si existe $\mathbf{a} \in \Re^n$ tal que

$$\mathbf{c}' = \mathbf{a}'\mathbf{X}$$

Lema: Supongamos que vale el modelo Ω. Sean $\psi = \mathbf{c}'\boldsymbol{\beta}$ una función estimable y $\mathcal{V}_r$ el espacio generado por las columnas de $\mathbf{X}$ ($r = rg(\mathbf{X}) \leq p$). Luego, existe un único estimador lineal insesgado de ψ, digamos $\mathbf{a}^{*'}\mathbf{Y}$ con $\mathbf{a}^* \in \mathcal{V}_r$. Más aún, si $\mathbf{a}'\mathbf{Y}$ es un estimador insesgado de ψ, $\mathbf{a}^*$ es la proyección ortogonal de $\mathbf{a}$ sobre $\mathcal{V}_r$.

Teorema de Gauss–Markov:
Supongamos que vale el modelo $\Omega : E(\mathbf{Y}) = \mathbf{X}\beta \quad \Sigma_{\mathbf{Y}} = \sigma^2\mathbf{I}$.
Toda función estimable $\psi = \mathbf{c}'\boldsymbol{\beta}$ tiene un único estimador $\widehat{\psi}$ lineal insesgado de míinima varianza (BLUE). Este estimador $\widehat{\psi}$ se puede obtener reemplazando a $\boldsymbol{\beta}$ en $\mathbf{c}'\boldsymbol{\beta}$ por $\widehat{\boldsymbol{\beta}}$, el estimador de mínimos cuadrados.

Definición: Dada una función estimable ψ su único estimador lineal insesgado de mínima varianza $\widehat{\psi}$, cuya existencia y cálculo están dados por el Teorema de Gauss–Markov, es el estimador de mínimos cuadrados de ψ.

Tenemos el siguiente resultado:

Corolario: Si $\{\psi_1, \ldots, \psi_q\}$ son q funciones estimables toda combinación lineal $\Psi = \Sigma_{i=1}^{q} h_i\psi_i$ es estimable y su estimador de mínimos cuadrado está dado por $\Sigma_{i=1}^{q} h_i\widehat{\psi_i}$.