Regresión Lineal Múltiple - Semana 05

Johnatan Cardona Jiménez
jcardonj@unal.edu.co
Profesor Asistente - Departamento de Estadística
Universidad Nacional de Colombia, Sede Medellín

Semestre 02-2024

Regresión lineal múltiple

- En análisis de regresión usualmente se utiliza más de una variable predictora para modelar el valor de una variable respuesta de interés. Si la relación funcional entre la respuesta Y y las variables predictoras es lineal en los parámetros, se llega al caso de la **regresión lineal múltiple** (RLM).
- En el ajuste y análisis de este modelo se obtienen resultados que son extensiones de los que se obtuvieron en regresión lineal simple.
- A continuación se introducen algunas nociones preliminares relacionadas con matrices y vectores aleatorios, las cuales permite un manejo matemático simplificado del modelo de RLM.

Nociones Preliminares relacionados con Vectores de Variables Aleatorias

Para facilitar la notación y el desarrollo de algunas pruebas se utiliza con frecuencia una escritura del modelo en forma matricial, la cual requiere establecer algunas definiciones que se presentan a continuación.

Vectores Aleatorios

Sean $y_1, y_2, \ldots, y_n$ variables aleatorias con medias $\mu_1, \mu_2, \ldots, \mu_n$ y varianzas $\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2$, respectivamente.

Sean $\sigma_{ij} = \text{Cov}\left[y_i, y_j\right] = E\left[\left(y_i - \mu_i\right)\left(y_j - \mu_j\right)\right]$ la covarianza entre las variables y_i e y_j, con $i, j = 1, 2, \ldots, n$.

Se define el vector $\underline{\boldsymbol{y}} = [y_1, y_2, \ldots, y_n]'$, el cual se dice es un **vector aleatorio** con **media o vector de medias** y **varianza o matriz de varianzas-covarianzas** dadas por: $\underline{\boldsymbol{\mu}}_{\underline{\boldsymbol{y}}} = E\left[\underline{\boldsymbol{y}}\right]$ y $Var(\underline{\boldsymbol{y}}) = \boldsymbol{\Sigma}_{\underline{\boldsymbol{y}}}$:

$$\underline{\boldsymbol{\mu}}_{\underline{\boldsymbol{y}}} = \begin{bmatrix} E\left[y_1\right] \\ \vdots \\ E\left[y_n\right] \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix} \quad \text{y} \quad \boldsymbol{\Sigma}_{\underline{\boldsymbol{y}}} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn} \end{bmatrix}$$

Observe que en la matriz $\boldsymbol{\Sigma}_{\underline{\boldsymbol{y}}}$ los elementos:

- $\sigma_{ij} = \sigma_{ji}$, es decir, la matriz $\boldsymbol{\Sigma}_{\underline{\boldsymbol{y}}}$ es una matriz simétrica.
- $\sigma_{ii} = \sigma_i^2$, es decir, los elementos de la diagonal principal de $\boldsymbol{\Sigma}_{\underline{\boldsymbol{y}}}$ corresponden a las varianzas de las variables y_i.

Por lo tanto se puede escribir,

$$\boldsymbol{\Sigma}_{\underline{\boldsymbol{y}}} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{12} & \sigma_2^2 & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1n} & \sigma_{2n} & \cdots & \sigma_n^2 \end{bmatrix}$$

En resumen, la matriz de varianzas-covarianzas asociada a un vector aleatorio es una matriz con las siguiestes características:

- Cuadrada y simétrica, de orden igual al tamaño del vector aleatorio.
- La diagonal principal contiene las varianzas asociadas a cada elemento del vector aleatorio.
- Por fuera de la diagonal están las covarianzas entre los pares de elementos del vector aleatorio.

Algunas propiedades del valor esperado y la varianza de un vector aleatorio

Sea $\underline{\boldsymbol{y}}$ un vector aleatorio $n \times 1$ con media (vector de medias) $\underline{\boldsymbol{\mu}}_{\underline{\boldsymbol{y}}}$ y varianza (matriz de varianzas-covarianzas) $\boldsymbol{\Sigma}_{\underline{\boldsymbol{y}}}$, como fueron definidos antes.

Sean $\boldsymbol{A}$ una matriz $m \times n$ de constantes y $\underline{\boldsymbol{b}}$ un vector $m \times 1$ de constantes, entonces:

1. $E\left[\boldsymbol{A}\underline{\boldsymbol{y}}\right] = \boldsymbol{A}E\left[\underline{\boldsymbol{y}}\right] = \boldsymbol{A}\mu_{\underline{\boldsymbol{y}}}$.
2. $E\left[\boldsymbol{A}\underline{\boldsymbol{y}} + \underline{\boldsymbol{b}}\right] = E\left[\boldsymbol{A}\underline{\boldsymbol{y}}\right] + E\left[\underline{\boldsymbol{b}}\right] = \boldsymbol{A}\mu_{\underline{\boldsymbol{y}}} + \underline{\boldsymbol{b}}$.
3. $Var\left[\boldsymbol{A}\underline{\boldsymbol{y}}\right] = \boldsymbol{A}\ Var\left[\underline{\boldsymbol{y}}\right]\boldsymbol{A}' = \boldsymbol{A}\ \boldsymbol{\Sigma}_{\underline{\boldsymbol{y}}}\ \boldsymbol{A}'$.
4. $Var\left[\boldsymbol{A}\underline{\boldsymbol{y}} + \underline{\boldsymbol{b}}\right] = Var\left[\boldsymbol{A}\underline{\boldsymbol{y}}\right] + Var\left[\underline{\boldsymbol{b}}\right] = \boldsymbol{A}\ \boldsymbol{\Sigma}_{\underline{\boldsymbol{y}}}\ \boldsymbol{A}' + \boldsymbol{O} = \boldsymbol{A}\ \boldsymbol{\Sigma}_{\underline{\boldsymbol{y}}}\ \boldsymbol{A}'$.

Observe que se conservan las propiedades de la esperanza y la varianza del caso univariado, por ejemplo.

- $E\left[\underline{\boldsymbol{b}}\right] = \underline{\boldsymbol{b}}$, es decir, la esperanza de un vector constante es el mismo vector constante.
- $Var\left[\underline{\boldsymbol{b}}\right] = \boldsymbol{O}$, donde $\boldsymbol{O}$ es una matriz nula (cuyos elementos son todos cero) de orden $m \times m$. Así, la varianza de un vector constante es una matriz cuadrada nula de orden correspondiente al número de elementos del vector constante.
- $Var\left[\boldsymbol{A}\underline{\boldsymbol{y}}\right] = \boldsymbol{A}\,\boldsymbol{\Sigma}_{\underline{\boldsymbol{y}}}\,\boldsymbol{A}'$, es decir, la varianza de un vector aleatorio por una matriz constante resulta en una forma cuadrática de la matriz constante que involucra a la varianza del vector aleatorio.

Algunas definiciones básicas en teoría matricial

Sean $\boldsymbol{A}$ y $\boldsymbol{B}$ matrices de constantes de orden $n \times n$ y $m \times n$ respectivamente.

Sean $\underline{\boldsymbol{x}} = [x_1, \ldots, x_n]'$ un vector de variables de orden $n \times 1$; $\underline{\boldsymbol{a}}$ un vector de constantes de orden $n \times 1$; y $\boldsymbol{I}$ la matriz identidad de orden n.

Entonces:

1. $(\boldsymbol{BA})' = \boldsymbol{A}'\boldsymbol{B}'$, la traspuesta de un producto es igual al producto invertido de las traspuestas.
2. $\boldsymbol{A}$ es simétrica si $\boldsymbol{A}' = \boldsymbol{A}$.
3. $\boldsymbol{A}$ es idempotente si $\boldsymbol{AA} = \boldsymbol{A}^2 = \boldsymbol{A}$.

4. Si $\boldsymbol{A}$ es simétrica e idempotente, entonces $(\boldsymbol{I} - \boldsymbol{A})$ también es simétrica e idempotente.
5. **Forma cuadrática:** la función $\underline{\boldsymbol{x}}'\boldsymbol{A}\underline{\boldsymbol{x}} = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_ix_j$ se le llama forma cuadrática de $\underline{\boldsymbol{x}}$, donde a_{ij} es la ij-ésima componente de la matriz $\boldsymbol{A}$.
6. **Matriz definida positiva y semidefinida positiva:** la matriz $\boldsymbol{A}$ se dice que es:
   a) Definida positiva, si $\underline{\boldsymbol{x}}'\boldsymbol{A}\underline{\boldsymbol{x}} > 0, \ \forall \underline{\boldsymbol{x}}$.
   b) Semidefinida positiva si $\underline{\boldsymbol{x}}'\boldsymbol{A}\underline{\boldsymbol{x}} \geq 0, \ \forall \underline{\boldsymbol{x}}$.

Algunas propiedades de derivadas Vectoriales o Matriciales

Sean $\boldsymbol{A}$ una matriz de constantes de orden $n \times n$; $\underline{\boldsymbol{x}} = [x_1, \ldots, x_n]'$ un vector de variables de orden $n \times 1$; y $\underline{\boldsymbol{a}}$ un vector de constantes de orden $n \times 1$.

Entonces:

1. $\dfrac{\partial\left(\underline{\boldsymbol{a}}'\underline{\boldsymbol{x}}\right)}{\partial\underline{\boldsymbol{x}}} = \dfrac{\partial\left(\underline{\boldsymbol{x}}'\underline{\boldsymbol{a}}\right)}{\partial\underline{\boldsymbol{x}}} = \underline{\boldsymbol{a}}.$
2. $\dfrac{\partial\left(\underline{\boldsymbol{x}}'\underline{\boldsymbol{x}}\right)}{\partial\underline{\boldsymbol{x}}} = 2\underline{\boldsymbol{x}}.$
3. $\dfrac{\partial\left(\underline{\boldsymbol{x}}'\boldsymbol{A}\underline{\boldsymbol{x}}\right)}{\partial\underline{\boldsymbol{x}}} = \boldsymbol{A}\underline{\boldsymbol{x}} + \boldsymbol{A}'\underline{\boldsymbol{x}}$, pero si $\boldsymbol{A}$ es simétrica, entonces $\dfrac{\partial\left(\underline{\boldsymbol{x}}'\boldsymbol{A}\underline{\boldsymbol{x}}\right)}{\partial\underline{\boldsymbol{x}}} = 2\boldsymbol{A}\underline{\boldsymbol{x}}.$

Definición del Modelo de Regresión Lineal Múltiple (RLM)

- Considere el caso en el cual se desea modelar la variabilidad total de una variable respuesta de interés, en función de relaciones lineales con dos o más variables predictoras, formuladas simultáneamente en un único modelo.
- Suponemos en principio que las variables predictoras guardan poca asociación lineal entre sí, es decir, cada variable predictora aporta información independiente de las demás predictoras presentes en el modelo (hasta cierto grado, la información aportada por cada una no es redundante).

La ecuación del modelo de regresión en este caso es:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik} + \varepsilon_i, \quad i = 1, 2, \ldots, n.$$

Este modelo es de primer orden ya que no se presentan efectos de interacción entre las k variables predictoras y tampoco se tienen términos no lineales en alguna de las variables explicativas. En este modelo se tiene que:

- $\beta_0, \beta_1, \ldots, \beta_k$: son los p parámetros del modelo (uno por cada variable predictora más uno por el intercepto, esto es $p = k + 1$).
- $X_{i1}, X_{i2}, \ldots, X_{ik}$, son los valores en la i-ésima observación muestral, de las k variables predictoras consideradas en el modelo.
- $\varepsilon_i \overset{\text{iid}}{\sim} N\left(0, \sigma^2\right), \quad i = 1, 2, \ldots, n.$

Estadísticamente, se establece que la respuesta media está dada por:

$$E\left(Y|X_1, X_2, \ldots, X_k\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k,$$

la cual representa un hiperplano en un espacio de dimensión $k+1$, llamado ***superficie de regresión o superficie de respuesta***.

Similar al modelo de regresión lineal simple, bajo los supuestos de normalidad, independencia y varianza constante de los errores, se tiene que:

$$Y_i|X_{i1}, \ldots, X_{ik} \overset{\text{ind}}{\sim} N\left(\beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik}\ ,\ \sigma^2\right), \quad i = 1, 2, \ldots, n.$$

Significado de los Coeficientes de Regresión en el Modelo Lineal Múltiple

- El parámetro β_0, el intercepto Y del plano, representará la respuesta media de Y cuando en el conjunto de observaciones se incluye la coordenada

$$(X_1, X_2, \ldots, X_k) = (0, 0, \ldots, 0),$$

  de lo contrario si tal coordenada no es observada o no está incluida en el rango experimental, entonces β_0 no será interpretable.
- Los parámetros $\beta_j, j = 1, 2, \ldots, k$, indican el cambio en la respuesta media de Y por unidad de incremento en la respectiva variable X_j, cuando las demás predictoras permanecen constantes (sin importar en qué nivel son fijadas estas últimas).

Como los efectos de una predictora sobre la respuesta media no dependen de los valores aquiridos por las demás, tales efectos son denominados *efectos aditivos*. Los parámetros β_j, son también llamados *coeficientes de regresión parcial* porque reflejan el efecto parcial de una variable predictora sobre la respuesta media en presencia de las demás predictoras que aparecen en el modelo.

NOTA: El término *modelo lineal* significa que el modelo es lineal en los parámetros, lo cual no hace referencia a la forma de la superficie de respuesta.

Tipos de Variables y de Efectos en los Modelos

Las variables predictoras pueden ser:

- Cuantitativas, caso en el cual se supone se miden sin error (o el error es despreciable).
- Cualitativas o categóricas, en este caso su manejo en el modelo se realiza a través de la definición de variables indicadoras, las cuales toman valores de 0 ó 1.

En general, una variable cualitativa con c clases se representa mediante $c-1$ variables indicadoras, puesto que cuando en una observación dada, todas las $c-1$ primeras indicadoras son iguales a cero, entonces la variable cualitativa se haya en su última clase.

Tipos de Variables y de Efectos en los Modelos

Por ejemplo, suponga que en un modelo de regresión para el gasto mensual por familia en actividades recreativas, se tiene entre las variables predictoras el estrato socioeconómico, definido en cinco niveles, luego, basta definir las primeras cuatro indicadoras de la siguiente forma:

$$I_1 = \begin{cases} 1 & \text{familia en estrato 1} \\ 0 & \text{otro caso} \end{cases} \qquad I_2 = \begin{cases} 1 & \text{familia en estrato 2} \\ 0 & \text{otro caso} \end{cases}$$

$$I_3 = \begin{cases} 1 & \text{familia en estrato 3} \\ 0 & \text{otro caso} \end{cases} \qquad I_4 = \begin{cases} 1 & \text{familia en estrato 4} \\ 0 & \text{otro caso} \end{cases}$$

Tipos de Variables y de Efectos en los Modelos

- En el caso con variables predictoras cuantitativas, existe la llamada ***regresión polinomial*** en la que se utilizan términos cuadráticos y de orden superior de estas variables, como en los diseños experimentales para optimización de procesos mediante la metodología de superficie de respuesta.
- A pesar de la naturaleza no lineal de tales superficies de respuesta, estos modelos hacen parte del modelo de regresión lineal múltiple.
- Algunos modelos pueden usar funciones de respuesta curvilíneas, en los cuales se utilizan variables transformadas de forma compleja, para linealización del modelo.

- Otros modelos pueden incluir *efectos de interacción*, es decir cuando los efectos de una variable predictora depende de los niveles de otras variables predictoras incluidas en el modelo.
- Por ejemplo, suponga un modelo de regresión con las variables predictoras X_1 y X_2, que incluye tanto los efectos principales como el efecto de interacción de estas dos variables. Este modelo corresponde a:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_1 X_2 + \varepsilon_i.$$

- El término de interacción es representado por $\beta_3 X_1 X_2$. Para expresar el anterior modelo en términos del modelo lineal múltiple, definimos simplemente $X_3 = X_1 X_2$ y rescribimos el modelo como:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_3 + \varepsilon_i.$$

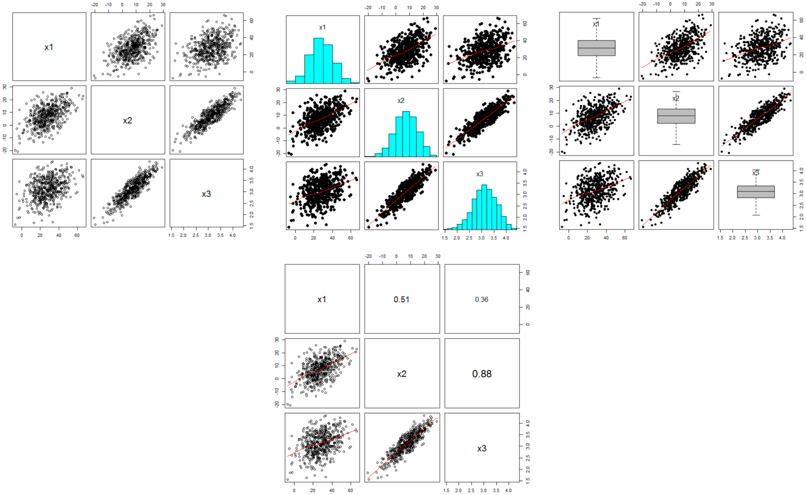
- En este último tipo de modelo los coeficientes de regresión de los términos individuales X_j ya no tienen el significado dado antes, debido a la presencia de la interacción, es decir, ya no representan el cambio en la respuesta media cuando se incrementa en una unidad la respectiva variable predictora, manteniendo constante a las demás.
- Para el ejemplo, puede mostrarse mediante derivación, que cuando X_1 se incrementa en una unidad mientras X_2 se deja fija, el cambio en la respuesta promedio es $\beta_1 + \beta_3 X_2$. Así, los efectos de una variable predictora sobre la respuesta promedio, dado el nivel fijo de la otra, dependen del nivel en que se halle esta última.
- Tenga presente que cualquier modelo que pueda rescribirse en términos del modelo lineal múltiple, puede trabajarse mediante las técnicas de estimación de mínimos cuadrados, de lo contrario, el modelo se considera no lineal y sólo pueden obtenerse estimaciones mediante métodos numéricos complejos.

Chequeo de Posibles Asociaciones

- Inicialmente, puede ser útil realizar chequeos gráficos de la naturaleza y la fuerza de las asociaciones entre las variables predictoras con la variable respuesta, y aún entre predictoras.
- Una matriz de gráficas de dispersión es la herramienta más útil para visualizar rápida y simultáneamente estas relaciones. Si las variables predictoras se asocian linealmente a la variable respuesta, los gráficos de dispersión respectivos deben presentar las nubes de puntos tendiendo a una línea recta. También se puede chequear si existen relaciones de tipo no lineal entre las distintas variables, y la presencia de observaciones atípicas.

- Por otra parte, se espera que entre las predictoras no existan relaciones lineales fuertes, pues de lo contrario, habría información que podría ser redundante en el modelo, y se tendría un *problema de multicolinealidad*, lo cuál se estudiará en mayor detalle más adelante en la asignatura.
- A veces es útil también acompañar este análisis gráfico con la matriz de correlaciones de las variables del modelo, la cual muestra los coeficientes de correlación entre la variable respuesta con cada una de las predictoras y también todas las correlaciones entre las predictoras.

En la siguiente figura se presentan algunas matrices de gráficos de dispersión para un conjunto de datos sobre tres variables.



Estimación por Mínimos Cuadrados de los Parámetros del Modelo

Note que cuando se tienen n observaciones para el modelo lineal

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik} + \varepsilon_i, \;\; i = 1, \ldots, n,$$

en realidad se tiene un sistema de n ecuaciones con $p = k + 1$ incógnitas correspondiendo al intercepto y los k coeficientes de regresión, donde los Y_i y las X_{ij} toman valores conocidos en cada caso.

$$\left\{ \begin{array}{ll} i = 1: & y_1 = \beta_0 + \beta_1 X_{11} + \beta_2 X_{12} + \cdots + \beta_k X_{1k} + \varepsilon_1 \\ i = 2: & y_2 = \beta_0 + \beta_1 X_{21} + \beta_2 X_{22} + \cdots + \beta_k X_{2k} + \varepsilon_2 \\ \vdots & \vdots \\ i = n: & y_n = \beta_0 + \beta_1 X_{n1} + \beta_2 X_{n2} + \cdots + \beta_k X_{nk} + \varepsilon_n \end{array} \right\}$$

Tal sistema expresado en forma matricial corresponde a:

$$\underline{\boldsymbol{y}} = \boldsymbol{X}\underline{\boldsymbol{\beta}} + \boldsymbol{\varepsilon}$$

donde:

$$\underline{\boldsymbol{y}} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}$$ es el vector $n \times 1$ de observaciones.

$$\underline{\boldsymbol{\beta}} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}$$ es el vector de $p \times 1$ parámetros.

$$\boldsymbol{X} = \begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1k} \\ 1 & X_{21} & X_{22} & \cdots & X_{2k} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_{nk} \end{bmatrix}$$ es una matriz $n \times p$ con los valores de predictoras.

$$\underline{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$ es el vector de errores aleatorios.

Recuerde que, los supuestos del modelo sobre los errores establecen que:

$$\varepsilon_i \overset{\text{iid}}{\sim} N\left(0, \sigma^2\right), \quad i = 1, 2, \ldots, n$$

Lo cual en forma matricial establece que el vector $\underline{\varepsilon}$ (de los errores aleatorios) es un vector aleatorio normal n-variado con valor esperado $E\left(\underline{\varepsilon}\right) = \mathbf{0}$ y matriz de varianzas covarianzas $Var\left(\underline{\varepsilon}\right) = \sigma^2 \boldsymbol{I}_n$, donde $\boldsymbol{I}_n$ es la matriz identidad de orden n.

Observe que $E\left(\underline{\varepsilon}\right) = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$ y $Var\left(\underline{\varepsilon}\right) = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix}$.

Por tanto el vector aleatorio $\underline{\boldsymbol{y}}$ tiene valor esperado $\boldsymbol{X}\underline{\beta}$ y la misma matriz de varianzas covarianzas de $\underline{\varepsilon}$.

Para la estimación por mínimos cuadrados se buscan los valores estimados de los parámetros tales que se minimice la suma de cuadrado del error:

$$S\left(\underline{\beta}\right) = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} \left(Y_i - \beta_0 - \beta_1 X_{i1} - \beta_2 X_{i2} - \cdots - \beta_k X_{ik}\right)^2 .$$

Matricialmente,

$$\begin{aligned} S\left(\underline{\beta}\right) &= \underline{\varepsilon}'\underline{\varepsilon} = \left(\underline{\boldsymbol{y}} - \boldsymbol{X}\underline{\beta}\right)'\left(\underline{\boldsymbol{y}} - \boldsymbol{X}\underline{\beta}\right) \\ &= \underline{\boldsymbol{y}}'\underline{\boldsymbol{y}} - \underline{\boldsymbol{y}}'\boldsymbol{X}\underline{\beta} - \underline{\beta}'\boldsymbol{X}'\underline{\boldsymbol{y}} + \underline{\beta}'\boldsymbol{X}'\boldsymbol{X}\underline{\beta} \\ &= \underline{\boldsymbol{y}}'\underline{\boldsymbol{y}} - 2\underline{\beta}'\boldsymbol{X}'\underline{\boldsymbol{y}} + \underline{\beta}'\boldsymbol{X}'\boldsymbol{X}\underline{\beta}, \end{aligned}$$

y el estimador de mínimos cuadrados se obtiene al resolver

$$\left.\frac{d\ S\left(\underline{\beta}\right)}{d\underline{\beta}} = 0\right|_{\underline{\beta}=\underline{\widehat{\beta}}} \iff -2\boldsymbol{X}d'\underline{\boldsymbol{y}} + 2\boldsymbol{X}'\boldsymbol{X}\underline{\widehat{\beta}} = 0$$

De donde las ecuaciones normales de mínimos cuadrados para el modelo de RLM son:

$$\boldsymbol{X}'\boldsymbol{X}\underline{\widehat{\boldsymbol{\beta}}} = \boldsymbol{X}'\underline{\boldsymbol{y}},$$

y el vector de los estimadores es $\underline{\widehat{\boldsymbol{\beta}}} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \ \boldsymbol{X}'\underline{\boldsymbol{y}}$, cuyos elementos corresponden a los estimadores de mínimos cuadrados para los parámetros del modelo de RLM. Esto es,

$$\underline{\widehat{\boldsymbol{\beta}}} = \left[\begin{array}{cccc} \widehat{\beta}_0 & \widehat{\beta}_1 & \cdots & \widehat{\beta}_k \end{array}\right]'$$

Luego, la ecuación de regresión ajustada igual a

$$\widehat{Y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 X_1 + \widehat{\beta}_2 X_2 + \cdots + \widehat{\beta}_k X_k = \underline{\boldsymbol{x}}_i \widehat{\boldsymbol{\beta}}.$$

donde, $\underline{\boldsymbol{x}}_i$ es la i-ésima fila de la matrix $\boldsymbol{X}$.

Algunas propiedades de los estimadores de los parámetros

- Los estimadores de mínimos cuadrados corresponden a los estimadores de máxima verosimilitud, bajo el modelo lineal normal.
- $\underline{\widehat{\beta}}$ es un estimador insesgado del vector de parámetros $\underline{\beta}$, es decir

$$E\left[\underline{\widehat{\beta}}\right] = E\left[\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\underline{\boldsymbol{y}}\right] = \underline{\beta}$$

En efecto, sea $\boldsymbol{A} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'$ una matriz $p \times n$ de valores fijos (constantes). Entonces,

$$E\left[\underline{\widehat{\beta}}\right] = E\left[\boldsymbol{A}\underline{\boldsymbol{y}}\right] = \boldsymbol{A}E\left[\underline{\boldsymbol{y}}\right] = \boldsymbol{A}\ \boldsymbol{X}\underline{\beta} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\ \boldsymbol{X}\underline{\beta} = \underline{\beta}$$

- La matriz de varianzas covarianzas de $\widehat{\underline{\beta}}$ es

$$Var\left(\widehat{\underline{\beta}}\right) = \sigma^2 \left(\boldsymbol{X'X}\right)^{-1}$$

En efecto,

$$\begin{aligned} Var\left[\widehat{\underline{\beta}}\right] &= Var\left[\boldsymbol{A}\underline{\boldsymbol{y}}\right] = \boldsymbol{A}Var\left[\underline{\boldsymbol{y}}\right]\boldsymbol{A}' = \boldsymbol{A}\,\sigma^2\boldsymbol{I}_n\,\boldsymbol{A}' \\ &= \sigma^2\left\{\left(\boldsymbol{X'X}\right)^{-1}\boldsymbol{X'}\boldsymbol{I}_n\left[\left(\boldsymbol{X'X}\right)^{-1}\boldsymbol{X'}\right]'\right\} \\ &= \sigma^2\left\{\left(\boldsymbol{X'X}\right)^{-1}\boldsymbol{X'}\,\boldsymbol{X}\left[\left(\boldsymbol{X'X}\right)^{-1}\right]'\right\} = \sigma^2\left(\boldsymbol{X'X}\right)^{-1}, \end{aligned}$$

La cual tiene en su diagonal principal a las varianzas de los estimadores de los parámetros, $Var\left(\widehat{\beta}_j\right), j = 0, 1, \ldots, k$, y por fuera de su diagonal principal a las covarianzas entre tales estimadores.

- Note que $\underline{\widehat{\beta}} = \boldsymbol{A}\underline{\boldsymbol{y}}$ implica que cada parámetro estimado es una combinación lineal de las observaciones, así que $\widehat{\beta}_j$ es una variable aleatoria con distribución normal (ya que los y_i's son normales).

En resumen, se tiene que:

$$\underline{\widehat{\beta}} \sim \boldsymbol{N}\left(\underline{\beta},\, \sigma^2 \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right),$$

y en el caso particular del estimador del j-ésimo parámetro del modelo se tiene que:

$$\widehat{\beta}_j \sim N\left(\beta_j,\, \sigma^2\, c_{jj}\right),\ j = 0, 1, \ldots, k.$$

Observe que en las expresiones anteriores σ^2 es desconocido, de manera que debe estimarse.

Un estimador de σ^2 insesgado se define como:

$$\hat{\sigma}^2 = \text{MSE} = \frac{\sum_{i=1}^{n} (y_i - \widehat{y}_i)^2}{n-p} = \frac{\text{SSE}}{n-p}.$$

Luego, una estimación de la matriz de varianzas-covarianzas es:

$$\widehat{Var}\left(\underline{\widehat{\boldsymbol{\beta}}}\right) = \hat{\sigma}^2 \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} = \text{MSE} \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1},$$

cuyos elementos en la diagonal principal corresponden a las estimaciones de las varianzas de los respectivos estimadores $\widehat{\beta}_j$, esto es,

$$\widehat{Var}\left(\widehat{\beta}_j\right) = \text{MSE}\, c_{jj},$$

donde c_{jj}, es el j-ésimo elemento de la diagonal de la matriz $\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}$.