CAPíTULO 2

Mínimos cuadrados ordinarios

En el Capítulo 1, hemos definido el modelo lineal general como una relación estadística lineal entre una variable dependiente y una o más variables explicativas, relación que podemos expresar de tres formas equivalentes:

1. $Y_i = \beta_1 + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i, \quad i = 1, \ldots, n$
2. $Y_i = \mathbf{x}_i' \boldsymbol{\beta} + u_i, \quad i = 1, \ldots, n$
3. $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$

en donde $\mathbf{x}_i = (1\, X_{2i} \, \ldots \, X_{ki})'$,

$$\mathbf{y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} 1 & X_{21} & \ldots & X_{k1} \\ 1 & X_{22} & \ldots & X_{k2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{2n} & \ldots & X_{kn} \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix} \quad \text{y} \quad \mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}.$$

La forma 1 y su versión compacta dada por la forma 2 son la representación escalar del modelo lineal general, y son útiles para introducir los conceptos básicos del método de estimación de mínimos cuadrados. Sin embargo, para profundizar en el estudio del análisis de regresión resulta más conveniente la forma 3 o forma matricial del modelo lineal general.

Este capítulo describe el álgebra de mínimos cuadrados, es decir, un método estadístico para obtener estimaciones de los parámetros desconocidos $\beta_1, \ldots, \beta_k$ a partir de un conjunto de observaciones sobre las variables Y, X_2, …, X_k. El método de estimación de mínimos cuadrados se presenta utilizando tanto la forma escalar como la forma matricial del modelo lineal general. Se deja como ejercicio para el alumno, el desarrollo de este procedimiento para la forma 2 del modelo.

2.1. Ecuaciones normales

2.1.1. Forma escalar. Si en la forma escalar del modelo lineal general reemplazamos los parámetros desconocidos $\beta_1, \ldots, \beta_k$ por las estimaciones $\hat{\beta}_1, \ldots, \hat{\beta}_k$, obtenemos el modelo estimado

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \cdots + \hat{\beta}_k X_{ki} + \hat{u}_i, \quad i = 1, \ldots, n$$

en donde $\hat{u}_i$ es una estimación del error u_i. Surgen aquí dos conceptos fundamentales.

Definición 2. *El valor ajustado $\hat{Y}_i$ es la combinación lineal*

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \cdots + \hat{\beta}_k X_{ki}, \quad i = 1, \ldots, n$$

Definición 3. *El residuo $\hat{u}_i$ es la diferencia entre el valor observado Y_i y el valor ajustado $\hat{Y}_i$,*

$$\hat{u}_i = Y_i - \hat{Y}_i = Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \cdots - \hat{\beta}_k X_{ki}, \quad i = 1, \ldots, n$$

Vemos que al estimar los parámetros del modelo lineal general descomponemos cada observación Y_i en la suma de dos componentes

$$Y_i = \hat{Y}_i + \hat{u}_i, \quad i = 1, \dots, n$$

en donde podemos interpretar el valor ajustado $\hat{Y}_i$ como la *predicción* de Y_i dada por el modelo estimado y el residuo $\hat{u}_i$ como el *error de predicción* asociado. Es claro que distintas estimaciones de los parámetros conducirán a distintos residuos o valores ajustados, siendo preferibles aquellas estimaciones que proporcionan un mayor número de residuos pequeños o, equivalentemente, un mayor número de valores ajustados muy próximos a los valores observados. Esta es la idea que subyace al método de estimación de mínimos cuadrados ordinarios.

Definición 4. *Las estimaciones de minimocuadráticas de los parámetros $\beta_1, \dots, \beta_k$ son los valores $\hat{\beta}_1, \dots, \hat{\beta}_k$ que minimizan la suma de cuadrados de los residuos*

$$Q = \sum_{i=1}^{n} \hat{u}_i^2 = \sum_{i=1}^{n} (\underbrace{Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \cdots - \hat{\beta}_k X_{ki}}_{\hat{u}_i})^2$$

Como Q es una función cuadrática de $\hat{\beta}_j$ $(j = 1, \dots, k)$, se trata de una función diferenciable. Los valores $\hat{\beta}_j$ $(j = 1, \dots, k)$ que minimizan Q son los que cumplen las condiciones de primer orden:

$$\begin{aligned}
\frac{\partial Q}{\partial \hat{\beta}_1} &= \sum_{i=1}^{n} 2(\underbrace{Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \cdots - \hat{\beta}_k X_{ki}}_{\hat{u}_i})(\underbrace{-1}_{\partial \hat{u}_i / \partial \hat{\beta}_1}) = 0 \\
\frac{\partial Q}{\partial \hat{\beta}_2} &= \sum_{i=1}^{n} 2(\underbrace{Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \cdots - \hat{\beta}_k X_{ki}}_{\hat{u}_i})(\underbrace{-X_{2i}}_{\partial \hat{u}_i / \partial \hat{\beta}_2}) = 0 \\
&\vdots \\
\frac{\partial Q}{\partial \hat{\beta}_k} &= \sum_{i=1}^{n} 2(\underbrace{Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \cdots - \hat{\beta}_k X_{ki}}_{\hat{u}_i})(\underbrace{-X_{ki}}_{\partial \hat{u}_i / \partial \hat{\beta}_k}) = 0
\end{aligned} \tag{2.1}$$

De las condiciones de primer orden obtenemos el sistema de k ecuaciones en k variables

$$\begin{aligned}
n\hat{\beta}_1 + \hat{\beta}_2 \sum_{i=1}^{n} X_{2i} + \cdots + \hat{\beta}_k \sum_{i=1}^{n} X_{ki} &= \sum_{i=1}^{n} Y_i \\
\hat{\beta}_1 \sum_{i=1}^{n} X_{2i} + \hat{\beta}_2 \sum_{i=1}^{n} X_{2i}^2 + \cdots + \hat{\beta}_k \sum_{i=1}^{n} X_{2i} X_{ki} &= \sum_{i=1}^{n} X_{2i} Y_i \\
&\vdots \\
\hat{\beta}_1 \sum_{i=1}^{n} X_{ki} + \hat{\beta}_2 \sum_{i=1}^{n} X_{ki} X_{2i} + \cdots + \hat{\beta}_k \sum_{i=1}^{n} X_{ki}^2 &= \sum_{i=1}^{n} X_{ki} Y_i
\end{aligned} \tag{2.2}$$

cuya solución permite encontrar las estimaciones de mínimos cuadrados $\hat{\beta}_j$ $(j = 1, \ldots, k)$. Cuando k es mayor que dos, es conveniente escribir (2.2) en forma matricial:

$$\underbrace{\begin{pmatrix} n & \sum_{i=1}^n X_{2i} & \ldots & \sum_{i=1}^n X_{ki} \\ \sum_{i=1}^n X_{2i} & \sum_{i=1}^n X_{2i}^2 & \ldots & \sum_{i=1}^n X_{2i}X_{ki} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^n X_{ki} & \sum_{i=1}^n X_{ki}X_{2i} & \ldots & \sum_{i=1}^n X_{ki}^2 \end{pmatrix}}_{\mathbf{X}'\mathbf{X}} \underbrace{\begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_k \end{pmatrix}}_{\hat{\boldsymbol{\beta}}} = \underbrace{\begin{pmatrix} \sum_{i=1}^n Y_i \\ \sum_{i=1}^n X_{2i}Y_i \\ \vdots \\ \sum_{i=1}^n X_{ki}Y_i \end{pmatrix}}_{\mathbf{X}'\mathbf{y}} \tag{2.3}$$

Vemos que el elemento genérico (i, j) de la matriz cuadrada $\mathbf{X}'\mathbf{X}$ de orden k es el producto escalar de las variables explicativas X_i y X_j, es decir, la suma de los productos de los valores de X_i por los correspondientes valores de X_j, $\sum_{h=1}^n X_{ih}X_{jh}$. De aquí, los elementos que se encuentran la diagonal principal (j, j) son sumas de cuadrados $\sum_{h=1}^n X_{jh}^2$ $(j = 1, \ldots, k)$, mientras que los que se encuentran en la primera fila o columna son sumas de observaciones: $\sum_{h=1}^n X_{1h}X_{jh} = \sum_{h=1}^n X_{jh}$ porque $X_{1h} = 1$ $(h = 1, \ldots, n)$. Es claro que el elemento (1,1) es el tamaño muestral porque $\sum_{h=1}^n X_{1h}^2 = \sum_{h=1}^n 1^2 = n$. Análogamente, el elemento i-ésimo del vector columna $\mathbf{X}'\mathbf{y}$ de orden k es el producto escalar de la variable explicativa X_i y la variable dependiente Y, $\sum_{h=1}^n X_{ih}Y_h$.

Definición 5. *Las ecuaciones (2.2) ó (2.3) se denominan <u>ecuaciones normales</u> y permiten obtener la expresión del estimador de mínimos cuadrados.*

Definición 6. *Los <u>estimadores minimocuadráticos</u> de los parámetros $\beta_1, \ldots, \beta_k$ del modelo lineal general son la solución del sistema (compatible determinado) de ecuaciones normales*

$$\underbrace{\begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_k \end{pmatrix}}_{\hat{\boldsymbol{\beta}}} = \underbrace{\begin{pmatrix} n & \sum_{i=1}^n X_{2i} & \ldots & \sum_{i=1}^n X_{ki} \\ \sum_{i=1}^n X_{2i} & \sum_{i=1}^n X_{2i}^2 & \ldots & \sum_{i=1}^n X_{2i}X_{ki} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^n X_{ki} & \sum_{i=1}^n X_{ki}X_{2i} & \ldots & \sum_{i=1}^n X_{ki}^2 \end{pmatrix}^{-1}}_{(\mathbf{X}'\mathbf{X})^{-1}} \underbrace{\begin{pmatrix} \sum_{i=1}^n Y_i \\ \sum_{i=1}^n X_{2i}Y_i \\ \vdots \\ \sum_{i=1}^n X_{ki}Y_i \end{pmatrix}}_{\mathbf{X}'\mathbf{y}} \tag{2.4}$$

cuyo cálculo requiere invertir una matriz no singular de orden $k \times k$.

Observación 2. Si la matriz $\mathbf{X}'\mathbf{X}$ es singular, entonces tenemos infinitas estimaciones que conducen a la misma suma de cuadrados de los residuos.

2.1.2. Forma matricial. Un desarrollo similar puede seguirse para el modelo en forma matricial. Reemplazando el vector de parámetros desconocidos $\boldsymbol{\beta}$ por el vector de estimaciones $\hat{\boldsymbol{\beta}}$, obtenemos el modelo estimado

$$\mathbf{y} = \mathbf{X}\hat{\boldsymbol{\beta}} + \hat{\mathbf{u}}$$

en donde $\hat{\mathbf{u}} = (\hat{u}_1 \ldots u_n)'$ es una estimación del vector de errores $\mathbf{u}$. Vemos que el modelo estimado descompone el vector de observaciones $\mathbf{y}$ en la suma del vector de valores ajustados $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$ y el vector de residuos $\hat{\mathbf{u}}$.

La suma de cuadrados de los residuos $\hat{\mathbf{u}}'\hat{\mathbf{u}}$ puede expresarse como

$$\begin{aligned}Q = \hat{\mathbf{u}}'\hat{\mathbf{u}} =& (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = (\mathbf{y}' - \hat{\boldsymbol{\beta}}'\mathbf{X}')(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})\\ =& \mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - \mathbf{y}'\mathbf{X}\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}}\\ =& \mathbf{y}'\mathbf{y} - 2\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} + \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}}\end{aligned}$$

en donde hemos usado la propiedad de la transposición de matrices, $(\mathbf{A} + \mathbf{B}\mathbf{C})' = \mathbf{A}' + \mathbf{C}'\mathbf{B}'$, y el resultado $\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} = \mathbf{y}'\mathbf{X}\hat{\boldsymbol{\beta}}$, que se cumple porque $\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y}$ es un escalar 1×1 y es igual a su traspuesta $\mathbf{y}'\mathbf{X}\hat{\boldsymbol{\beta}}$. Vemos que $\hat{\mathbf{u}}'\hat{\mathbf{u}}$ depende de una forma lineal y de una forma cuadrática en $\hat{\boldsymbol{\beta}}$.

Para hallar el mínimo de $\hat{\mathbf{u}}'\hat{\mathbf{u}}$ respecto de $\hat{\boldsymbol{\beta}}$ igualamos el vector de primeras derivadas al vector nulo

$$\frac{\partial(\hat{\mathbf{u}}'\hat{\mathbf{u}})}{\partial\hat{\boldsymbol{\beta}}} = -2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{0}_k$$

De aquí, obtenemos el sistema de ecuaciones normales dado en (2.3)

$$\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}$$

y la expresión del estimador de mínimos cuadrados ordinarios de la definición (6)

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y}$$

Para asegurarnos de que $\hat{\boldsymbol{\beta}}$ es un mínimo debemos comprobar que la matriz hessiana o matriz de segundas derivadas es definida positiva

$$\frac{\partial^2(\hat{\mathbf{u}}'\hat{\mathbf{u}})}{\partial\hat{\boldsymbol{\beta}}\partial\hat{\boldsymbol{\beta}}'} = \left[\frac{\partial^2(\hat{\mathbf{u}}'\hat{\mathbf{u}})}{\partial\hat{\beta}_i\partial\hat{\beta}_j}\right] = 2\mathbf{X}'\mathbf{X}$$

PROPOSICIÓN 1. *Bajo el supuesto de que* $\mathbf{X}'\mathbf{X}$ *es no singular, la matriz* $\mathbf{X}'\mathbf{X}$ *es una matriz definida positiva.*

DEMOSTRACIÓN. La matriz cuadrada $\mathbf{X}'\mathbf{X}$ de orden k es definida positiva si para cualquier vector columna $\mathbf{a}$ de orden k la forma cuadrática

$$\mathbf{a}'\mathbf{X}'\mathbf{X}\mathbf{a} > 0$$

Definiendo el vector columna $\mathbf{z} = \mathbf{X}\mathbf{a}$, tenemos

$$\mathbf{a}'\mathbf{X}'\mathbf{X}\mathbf{a} = \mathbf{z}'\mathbf{z} = \sum_{i=1}^{n} z_i^2 \geq 0$$

La igualdad surge cuando $\mathbf{z}$ es un vector nulo o, equivalentemente, cuando las columnas de $\mathbf{X}$ son linealmente dependientes. Si $\mathbf{X}'\mathbf{X}$ es una matriz no singular, $rang(\mathbf{X}'\mathbf{X}) = rang(\mathbf{X}) = k$, por lo que las columnas de $\mathbf{X}$ son linealmente independientes. □

2.2. Propiedades numéricas del método de mínimos cuadrados

La propiedad numérica fundamental del método de mínimos cuadrados se deriva directamente de las condiciones de primer order orden.

PROPOSICIÓN 2. *Los residuos y las variables explicativas son ortogonales*

$$\sum_{i=1}^{n} \hat{u}_i X_{ji} = 0, \quad j = 1, \ldots, k$$

Demostración. Ver las condiciones de primer orden (2.1). Alternativamente, las ecuaciones normales del modelo en forma matricial

$$\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{X}'\mathbf{y} = \mathbf{0}_k$$

pueden escribirse como

$$\mathbf{X}'\left[\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right] = \mathbf{X}'\hat{\mathbf{u}} = \mathbf{0}_k$$

cuyo elemento j-ésimo es el producto escalar de los residuos y la variable explicativa X_j. □

Corolario 1. *Si el modelo incluye término constante, la suma de los residuos es igual a cero.*

Demostración. Si el modelo incluye término constante, la primera variable asociada a β_1, X_{1i}, es una variable de unos, de modo que

$$\sum_{i=1}^{n} \hat{u}_i X_{1i} = \sum_{i=1}^{n} \hat{u}_i = 0$$

□

Corolario 2. *Los residuos $\hat{u}_i$ y los valores ajustados $\hat{Y}_i$ son ortogonales*

$$\sum_{i=1}^{n} \hat{Y}_i \hat{u}_i = 0$$

Demostración. Los valores ajustados son una combinación lineal de las variables explicativas. Como las variables explicativas y los residuos son otogonales también lo serán los valores ajustados y los residuos

$$\begin{aligned}\sum_{i=1}^{n} \hat{Y}_i \hat{u}_i =& \sum_{i=1}^{n} (\hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \cdots + \hat{\beta}_k X_{ki})\hat{u}_i \\ =& \hat{\beta}_1 \sum_{i=1}^{n} \hat{u}_i + \hat{\beta}_2 \sum_{i=1}^{n} \hat{u}_i X_{2i} + \cdots + \hat{\beta}_k \sum_{i=1}^{n} \hat{u}_i X_{ki} = 0\end{aligned}$$

Alternativamente, vemos que

$$\hat{\mathbf{y}}'\hat{\mathbf{u}} = \hat{\boldsymbol{\beta}}'\mathbf{X}'\hat{\mathbf{u}} = \mathbf{0}_k$$

□

Corolario 3. *Si el modelo incluye término constante, la media de las observaciones de la variable dependiente es igual a la media de los valores ajustados*

$$\frac{1}{n}\sum_{i=1}^{n} Y_i = \frac{1}{n}\sum_{i=1}^{n} \hat{Y}_i$$

Demostración. De la descomposición $Y_i = \hat{Y}_i + \hat{u}_i$, se cumple que

$$\frac{1}{n}\sum_{i=1}^{n} Y_i = \frac{1}{n}\sum_{i=1}^{n} (\hat{Y}_i + \hat{u}_i) = \frac{1}{n}\sum_{i=1}^{n} \hat{Y}_i + \frac{1}{n}\sum_{i=1}^{n} \hat{u}_i = \frac{1}{n}\sum_{i=1}^{n} \hat{Y}_i$$

□

Observación 3. *El corolario anterior indica que el hiperplano de regresión $\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \cdots + \hat{\beta}_k X_{ki}$ pasa por el punto $(\bar{Y}, \bar{X}_2, \ldots, \bar{X}_k)$.*

Observación 4. La propiedad numérica fundamental del método de mínimos cuadrados (los residuos y las variables explicativas son ortogonales) nos sugiere una regla práctica para especificar las ecuaciones normales. El punto de partida es el modelo estimado

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \cdots + \hat{\beta}_k X_{ki} + \hat{u}_i, \quad i = 1, \ldots, n$$

La j-ésima ecuación normal puede obtenerse multiplicando el modelo estimado por la variable j-ésima X_{ji}

$$X_{ji}Y_i = \hat{\beta}_1 X_{ji} + \hat{\beta}_2 X_{ji}X_{2i} + \cdots + \hat{\beta}_k X_{ji}X_{ki} + X_{ji}\hat{u}_i, \quad i = 1, \ldots, n$$

Sumando todas las observaciones, tenemos

$$\sum_{i=1}^{n} X_{ji}Y_i = \hat{\beta}_1 \sum_{i=1}^{n} X_{ji} + \hat{\beta}_2 \sum_{i=1}^{n} X_{ji}X_{2i} + \cdots + \hat{\beta}_k \sum_{i=1}^{n} X_{ji}X_{ki}, \quad j = 1, \ldots, k$$

porque $\sum_{i=1}^{n} X_{ji}\hat{u}_i$. Análogamente, premultiplicando el modelo estimado

$$\mathbf{y} = \mathbf{X}\hat{\boldsymbol{\beta}} + \hat{\mathbf{u}}$$

por la traspuesta de $\mathbf{X}$, tenemos

$$\mathbf{X}'\mathbf{y} = \mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}}$$

Ahora podemos encontrar fácilmente la expresión del estimador de mínimos cuadrados para la forma escalar compacta del modelo. Premultiplicando el modelo estimado

$$y_i = \mathbf{x}_i'\hat{\boldsymbol{\beta}} + u_i, \quad i = 1, \ldots, n$$

por el vector columna $\mathbf{x}_i$, y sumando para todas las observaciones, obtenemos

$$\sum_{i=1}^{n} \mathbf{x}_i y_i = \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i' \hat{\boldsymbol{\beta}} + \sum_{i=1}^{n} \mathbf{x}_i u_i, \quad i = 1, \ldots, n$$

en donde el elemento j-ésimo de $\sum_{i=1}^{n} \mathbf{x}_i u_i$ es $\sum_{i=1}^{n} X_{ji} u_i = 0$. De aquí, el estimador minimocuadrático puede expresarse como

$$\boxed{\hat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i' \right)^{-1} \sum_{i=1}^{n} \mathbf{x}_i y_i}$$

2.3. Regresión con datos centrados

Cuando el modelo a estimar incluye un término constante es conveniente usar datos centrados o en desviaciones respecto a la media. De esta forma, podemos estimar las $k-1$ pendientes $\beta_2, \ldots, \beta_k$ resolviendo un sistema de $k-1$ ecuaciones normales, estimaciones que usamos después para estimar la ordenada β_1. Hay que valorar que la inversión de una matriz de orden $k-1$ es siempre menos costosa que la de una matriz de orden k.

Partiendo del modelo estimado por mínimos cuadrados

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \cdots + \hat{\beta}_k X_{ki} + \hat{u}_i, \quad i = 1, \ldots, n \tag{2.5}$$

si sumamos todas las observaciones de Y y dividimos por n tenemos

$$\frac{1}{n}\sum_{i=1}^{n} Y_i = \frac{1}{n}\sum_{i=1}^{n} \hat{\beta}_1 + \hat{\beta}_2 \frac{1}{n}\sum_{i=1}^{n} X_{2i} + \cdots + \hat{\beta}_k \frac{1}{n}\sum_{i=1}^{n} X_{ki} + \frac{1}{n}\sum_{i=1}^{n} \hat{u}_i$$

y, dado que la suma de los residuos es cero,

$$\bar{Y} = \hat{\beta}_1 + \hat{\beta}_2 \bar{X}_2 + \cdots + \hat{\beta}_k \bar{X}_k \tag{2.6}$$

Restando (2.6) de (2.5), tenemos

$$Y_i - \bar{Y} = \hat{\beta}_2(X_{2i} - \bar{X}_2) + \cdots + \hat{\beta}_k(X_{ki} - \bar{X}_k) + \hat{u}_i, \quad i = 1, \ldots, n$$

que, para simplificar, escribimos como

$$y_i = \hat{\beta}_2 x_{2i} + \cdots + \hat{\beta}_k x_{ki} + \hat{u}_i, \quad i = 1, \ldots, n \tag{2.7}$$

en donde $y_i = Y_i - \bar{Y}$ y $x_{ji} = X_{ji} - \bar{X}_j$ son los valores de cada variable en desviaciones con respecto a su media. Observamos que desaparece el término constante, pero los parámetros estimados $\hat{\beta}_2, \ldots, \hat{\beta}_k$ y los residuos permanecen inalterados. Por tanto, la aplicación del método de mínimos cuadrados en la ecuación (2.7) proporciona los mismos resultados que en la ecuación (2.5). Los estimadores de mínimos cuadrados $\hat{\beta}_2, \ldots, \hat{\beta}_k$ pueden obtenerse utilizando datos en desviaciones con respecto a la media a partir de la fórmula

$$\begin{pmatrix} \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_k \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n x_{2i}^2 & \ldots & \sum_{i=1}^n x_{2i}x_{ki} \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^n x_{ki}x_{2i} & \ldots & \sum_{i=1}^n x_{ki}^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum_{i=1}^n x_{2i}y_i \\ \vdots \\ \sum_{i=1}^n x_{ki}y_i \end{pmatrix} \tag{2.8}$$

cuyo cálculo requiere invertir una matriz de menor orden $(k-1) \times (k-1)$. El término constante se obtiene de la ecuación (2.6) que relaciona las medias

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}_2 - \cdots - \hat{\beta}_k \bar{X}_k$$

2.4. Regresión con datos tipificados

Las pendientes estimadas del modelo lineal general miden la respuesta de la variable dependiente a cambios unitarios en las variables explicativas. El tamaño de estas estimaciones depende de las unidades en que se midan las variables. De aquí, para apreciar la importancia relativa de cada variable explicativa es conveniente usar datos tipicados.

Definición 7. *Una variable tipificada, normalizada o estandarizada es una transformación lineal de una variable observada que se obtiene restando de cada observación la media y diviendo la desviación resultante por la desviación típica. Para las variables del modelo lineal general, las variables tipificadas son*

$$y_i^* = \frac{(Y_i - \bar{Y})}{S_Y} \quad y \quad x_{ji}^* = \frac{(X_{ji} - \bar{X})}{S_{X_j}}$$

en donde

$$S_Y^2 = \frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{n} \quad y \quad S_{X_j}^2 = \frac{\sum_{i=1}^n (X_{ji} - \bar{X}_j)^2}{n}$$

son las varianzas muestrales de Y y X_j, respectivamente.

Observación 5. *Las variables tipificadas tienen media cero y varianza unitaria. Por ejemplo, la variable dependiente tipificada tiene media*

$$\frac{\sum_{i=1}^n y_i^*}{n} = \frac{\sum_{i=1}^n (Y_i - \bar{Y})/S_Y}{n} = \frac{\sum_{i=1}^n Y_i - n\bar{Y}}{S_Y n} = \frac{\bar{Y} - \bar{Y}}{S_Y} = 0$$

y varianza

$$\frac{\sum_{i=1}^{n}(y_i^* - 0)^2}{n} = \frac{\sum_{i=1}^{n}[(Y_i - \bar{Y})/S_Y]^2}{n} = \frac{S_Y^2}{S_Y^2} = 1$$

PROPOSICIÓN 3. *La regresión con datos tipificados es*

$$y_i^* = \hat{\beta}_2^* x_{2i} + \cdots + \hat{\beta}_k^* x_{ki}^* + \hat{u}_i^*, \quad i = 1, \ldots, n$$

en donde los nuevos coeficientes de regresión y residuos son

$$\hat{\beta}_j^* = \hat{\beta}_j \frac{S_{X_j}}{S_Y} \, (j = 1, \ldots, k) \quad y \quad \hat{u}_i^* = \hat{u}_i / S_Y \, (i = 1, \ldots, n)$$

DEMOSTRACIÓN. Dividiendo el modelo estimado con datos centrados

$$y_i = \hat{\beta}_2 x_{2i} + \cdots + \hat{\beta}_k x_{ki} + \hat{u}_i, \quad i = 1, \ldots, n$$

por S_y y multiplicando y diviendo cada variable explicativa por S_{X_j}, obtenemos

$$\underbrace{\frac{y_i}{S_y}}_{y_i^*} = \underbrace{\hat{\beta}_2 \frac{S_{X_2}}{S_y}}_{\hat{\beta}_2^*} \underbrace{\frac{x_{2i}}{S_{X_2}}}_{x_{2j}^*} + \cdots + \underbrace{\hat{\beta}_k \frac{S_{X_k}}{S_y}}_{\hat{\beta}_k^*} \underbrace{\frac{x_{ki}}{S_{X_k}}}_{x_{ki}^*} + \underbrace{\frac{\hat{u}_i}{S_y}}_{\hat{u}_i^*}, \quad i = 1, \ldots, n$$

□

PROPOSICIÓN 4. *Las pendientes estimadas en la regresión con datos tipificados miden la respuesta de la variable dependiente en términos de su desviación típica ante un cambio de una desviación típica en las variables explicativas.*

DEMOSTRACIÓN. Un cambio cambio unitario en la variable tipificada x_j^* es equivalente a un cambio de una desviación típica en la variable observada

$$\Delta x_{ji}^* = x_{ji}^* - x_{j,i-1}^* = \frac{X_{ji} - \bar{X}_j}{S_{X_j}} - \frac{X_{j,i-1} - \bar{X}_j}{S_{X_j}} = \frac{X_{ji} - X_{j,i-1}}{S_{X_j}} = \frac{\Delta X_{ji}}{S_{X_j}}$$

en donde Δx_{ji}^* será igual a uno cuando $\Delta X_{ji} = S_{X_j}$. A su vez, un cambio de una desviación típica en X_j implica la siguiente respuesta en Y

$$\Delta Y_i = Y_i - Y_{i-1} = \hat{\beta}_j S_{X_j}$$

o, en términos de su desviación típica,

$$\frac{\Delta Y_i}{S_y} = \hat{\beta}_j \frac{S_{X_j}}{S_Y} = \hat{\beta}_j^*$$

□

PROPOSICIÓN 5. *El coeficiente de determinación es invariante a transformaciones lineales de las variables del modelo y, en particular, a la tipificación de datos.*

DEMOSTRACIÓN. El coeficiente de determinación en el modelo con datos tipificados es

$$R^{*^2} = 1 - \frac{\sum_{i=1}^{n} \hat{u}_i^{*^2}}{\sum_{i=1}^{n} y_i^{*^2}} = 1 - \frac{\sum_{i=1}^{n} \hat{u}_i^2 / S_y^2}{n} = 1 - \frac{\sum_{i=1}^{n} \hat{u}_i^2}{\sum_{i=1}^{n} y_i^2} = R^2$$

□

2.5. Las matrices P y M

Las matrices $\mathbf{P}$ y $\mathbf{M}$ de orden $n \times n$

$$\mathbf{P} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \quad \text{y} \quad \mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$$

son dos matrices fundamentales en el análisis de regresión. Estas matrices están asociadas a los vectores de valores ajustados y de residuos, respectivamente. Sustituyendo $\hat{\boldsymbol{\beta}}$ por su expresión en $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$ tenemos

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

Análogamente,

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

y sacando $\mathbf{y}$ como factor común por la derecha

$$\hat{\mathbf{u}} = \left[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right]\mathbf{y} = \mathbf{M}\mathbf{y}$$

Definición 8. *La <u>transformación lineal</u> de un vector* $\mathbf{y}$ *a un vector* $\mathbf{z}$*, ambos de orden* $n \times 1$*, se define como* $\mathbf{z} = \mathbf{A}\mathbf{y}$*, en donde* $\mathbf{A}$ *es una matriz cuadrada* $n \times n$ *de coeficientes.*

Definición 9. *Una transformación lineal* $\mathbf{z} = \mathbf{A}\mathbf{y}$ *es una <u>proyección</u>, si la matriz* $\mathbf{A}$ *es simétrica* ($\mathbf{A} = \mathbf{A}'$) *e idempotente* ($\mathbf{A} = \mathbf{A}\mathbf{A}$).

Observación 6. *Si* $\mathbf{A}$ *es una matriz de proyección (simétrica e idempotente),* $\mathbf{I} - \mathbf{A}$ *es también una matriz de proyección.*

Proposición 6. *Las matrices* $\mathbf{P}$ *y* $\mathbf{M}$ *son matrices de proyección.*

Demostración.

1. $\mathbf{P}$ es una matriz simétrica $\mathbf{P} = \mathbf{P}'$

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

   porque

$$(\mathbf{A}(\mathbf{B}\mathbf{C})^{-1}\mathbf{D})' = \mathbf{D}'(\mathbf{C}'\mathbf{B}')^{-1}\mathbf{A}'$$

2. $\mathbf{P}$ es una matriz idempotente $\mathbf{P} = \mathbf{P}\mathbf{P}$

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

3. $\mathbf{M}$ es una matriz simétrica

$$\mathbf{M}' = (\mathbf{I} - \mathbf{P})' = \mathbf{I}' - \mathbf{P}' = \mathbf{I} - \mathbf{P} = \mathbf{M}$$

4. $\mathbf{M}$ es una matriz idempotente

$$\mathbf{M}\mathbf{M} = (\mathbf{I} - \mathbf{P})(\mathbf{I} - \mathbf{P}) = \mathbf{I} - \mathbf{P} - \mathbf{P} + \mathbf{P}\mathbf{P} = \mathbf{I} - \mathbf{P} = \mathbf{M}$$

$\square$

Corolario 4. *El producto de la matriz* $\mathbf{M}$ *de dimensión* $n \times n$ *por el vector columna de residuos* $\hat{\mathbf{u}}$ *es igual a* $\hat{\mathbf{u}}$*, esto es,* $\mathbf{M}\hat{\mathbf{u}} = \hat{\mathbf{u}}$.

Demostración.

$$\mathbf{M}\hat{\mathbf{u}} = \mathbf{M}\mathbf{M}\mathbf{y} = \mathbf{M}\mathbf{y} = \hat{\mathbf{u}}$$

□

PROPOSICIÓN 7. *Las matrices* $\mathbf{P}$ *y* $\mathbf{M}$ *son ortogonales,* $\mathbf{PM} = \mathbf{0}$.

COROLARIO 5. *El producto de* $\mathbf{M}$ *por* $\mathbf{X}$ *es una matriz nula de orden* $n \times k$.

DEMOSTRACIÓN.

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

□

PROPOSICIÓN 8. *El rango de* $\mathbf{P}$ *es* k *y el de* $\mathbf{M}$ $n - k$.

DEMOSTRACIÓN. El rango de una matriz idempotente es igual a su traza. De aquí,

$$\begin{aligned} tr(\mathbf{P}) =& tr[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'] = tr[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}] = tr(\mathbf{I}_k) = k \\ tr(\mathbf{M}) =& tr(\mathbf{I}_n - \mathbf{P}) = tr(\mathbf{I}_n) - tr(\mathbf{P}) = n - k \end{aligned}$$

en donde se han utilizado las siguientes propiedades de la traza: $trAB = trBA$ y $tr(A + B) = trA + trB$. □

De todo lo anterior deducimos que el producto de la matriz $\mathbf{P}$ de dimensión $n \times n$ por un vector columna $\mathbf{z}$ de dimensión $n \times 1$ es el vector de valores ajustados $\hat{\mathbf{z}}$, en la regresión de $\mathbf{z}$ sobre $\mathbf{X}$. Para destacar este resultado nos referiremos a $\mathbf{P}$ como la matriz generadora de valores ajustados. Del mismo modo, el producto de la matriz $\mathbf{M}$ de dimensión $n \times n$ por un vector columna $\mathbf{z}$ de dimensión $n \times 1$ es el vector de residuos $\hat{\mathbf{u}}$, en la regresión de $\mathbf{z}$ sobre $\mathbf{X}$; y nos referiremos a $\mathbf{M}$ como la matriz generadora de residuos.

2.6. Regresión particionada

El modelo lineal general estimado por mínimos cuadrados

$$\mathbf{y} = \mathbf{X}\hat{\boldsymbol{\beta}} + \hat{\mathbf{u}}$$

puede escribirse como

$$\mathbf{y} = \mathbf{X}_r\hat{\boldsymbol{\beta}}_r + \mathbf{X}_s\hat{\boldsymbol{\beta}}_s + \hat{\mathbf{u}} \tag{2.9}$$

en donde $\mathbf{X}_r$ es una submatriz $n \times r$ que contiene las r primeras columnas de la matriz $\mathbf{X}$ y $\mathbf{X}_s$ es una submatriz $n \times s$ que contiene las $s = k - r$ columnas restantes de $\mathbf{X}$; análogamente, $\hat{\boldsymbol{\beta}}_r$ es un subvector columna $r \times 1$ que contiene los r primeros parámetros del vector $\hat{\boldsymbol{\beta}}$ y $\hat{\boldsymbol{\beta}}_s$ es un subvector columna $s \times 1$ que contiene los $s = k - r$ restantes parámetros del vector $\hat{\boldsymbol{\beta}}$.

Sea la matriz

$$\mathbf{M}_s = \mathbf{I}_n - \mathbf{X}_s(\mathbf{X}_s'\mathbf{X}_s)^{-1}\mathbf{X}_s'$$

Por analogía con la matriz $\mathbf{M}$ tenemos

1. $\mathbf{M}_s\mathbf{X}_s = \mathbf{0}$ es la matriz nula de orden $n \times s$,
2. $\mathbf{M}_s\mathbf{y}$ es el vector de residuos $\hat{\mathbf{u}}_s$ de orden $n \times 1$ en la regresión de $\mathbf{y}$ sobre $\mathbf{X}_s$

$$\mathbf{y} = \mathbf{X}_s\mathbf{b}_s + \hat{\mathbf{u}}_s$$

con $\mathbf{b}_s = (\mathbf{X}_s'\mathbf{X}_s)^{-1}\mathbf{X}_s'\mathbf{y}$,

3. $\mathbf{M}_s\mathbf{X}_r$ es una matriz $n \times k$ cuya columna j son los residuos en la regresión de la variable X_j sobre $\mathbf{X}_s$,
4. $\mathbf{M}_s\hat{\mathbf{u}} = \hat{\mathbf{u}}$.

Premultiplicando la ecuación (2.9) por la matriz $\mathbf{M}_s$ tenemos

$$\mathbf{M}_s\mathbf{y} = \mathbf{M}_s\mathbf{X}_r\hat{\boldsymbol{\beta}}_r + \mathbf{M}_s\mathbf{X}_s\hat{\boldsymbol{\beta}}_s + \mathbf{M}_s\hat{\mathbf{u}}$$

o bien

$$\mathbf{M}_s\mathbf{y} = \mathbf{M}_s\mathbf{X}_r\hat{\boldsymbol{\beta}}_r + \hat{\mathbf{u}}$$

Aplicando la fórmula del estimador de mínimos cuadrados a este modelo de regresión, tenemos

$$\hat{\boldsymbol{\beta}}_r = (\mathbf{X}_r'\mathbf{M}_s'\mathbf{M}_s\mathbf{X}_r)^{-1}\mathbf{X}_r'\mathbf{M}_s'\mathbf{M}_s\mathbf{y} = (\mathbf{X}_r'\mathbf{M}_s\mathbf{X}_r)^{-1}\mathbf{X}_r'\mathbf{M}_s\mathbf{y}$$

Teorema 1. *Frisch-Waugh-Lovel. En el modelo de regresión particionado (2.9) los estimadores de mínimos cuadrados de $\boldsymbol{\beta}_r$ y $\boldsymbol{\beta}_s$ son*

$$\hat{\boldsymbol{\beta}}_r =(\mathbf{X}_r'\mathbf{M}_s\mathbf{X}_r)^{-1}\mathbf{X}_r'\mathbf{M}_s\mathbf{y}$$
$$\hat{\boldsymbol{\beta}}_s =(\mathbf{X}_s'\mathbf{M}_r\mathbf{X}_s)^{-1}\mathbf{X}_s'\mathbf{M}_r\mathbf{y}$$

Corolario 6. *Modelo en desviaciones respecto a la media. Si $\mathbf{X}_r = \mathbf{i} = (1\,1\,\ldots\,1)'$, entonces*

$$\mathbf{y} = \mathbf{i}\hat{\beta}_1 + \mathbf{X}_s\hat{\boldsymbol{\beta}}_s + \hat{\mathbf{u}}$$

y premultiplicando por $\mathbf{M}_i = \mathbf{I}_n - \mathbf{i}(\mathbf{i}'\mathbf{i})^{-1}\mathbf{i}'$ tenemos

$$\mathbf{M}_i\mathbf{y} = \mathbf{M}_i\mathbf{X}_s\hat{\boldsymbol{\beta}}_s + \hat{\mathbf{u}}$$

en donde

$$\mathbf{M}_i\mathbf{y} = \begin{pmatrix} Y_1 - \bar{Y} \\ Y_2 - \bar{Y} \\ \vdots \\ Y_n - \bar{Y} \end{pmatrix} \quad y \quad \mathbf{M}_i\mathbf{X}_s = \begin{pmatrix} X_{21} - \bar{X}_2 & \ldots & X_{k1} - \bar{X}_k \\ X_{22} - \bar{X}_2 & \ldots & X_{k2} - \bar{X}_k \\ \vdots & \ldots & \vdots \\ X_{2n} - \bar{X}_2 & \ldots & X_{kn} - \bar{X}_k \end{pmatrix}$$

contienen los datos en desviaciones. Aplicando la fórmula del estimador de mínimos cuadrados ordinarios al modelo en desviaciones

$$\hat{\boldsymbol{\beta}}_s = (\mathbf{X}_s'\mathbf{M}_i\mathbf{X}_s)^{-1}\mathbf{X}_s'\mathbf{M}_i\mathbf{y}$$

que es la forma simplificada de la ecuación (2.8).

Proposición 9. *La matriz $\mathbf{M}_i$ transforma un vector de observaciones $\mathbf{z} = [Z_i]$ de orden $n \times 1$ en un vector que contiene las observaciones en desviaciones con respecto a la media $\mathbf{M}_i\mathbf{z} = [Z_i - \bar{Z}]$*

Demostración. El producto $\mathbf{M}_i\mathbf{z}$ se interpeta como los residuos en la regresión de $\mathbf{z}$ sobre el vector $\mathbf{i}$

$$\mathbf{z} = \hat{\beta}_1\mathbf{i} + \hat{\mathbf{u}}$$

Pero $\hat{\beta}_1 = (\mathbf{i}'\mathbf{i})^{-1}\mathbf{i}'\mathbf{z} = \bar{Z}$, por tanto,

$$\hat{\mathbf{u}} = \mathbf{z} - \bar{Z}\mathbf{i} = \begin{pmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_n \end{pmatrix} - \begin{pmatrix} \bar{Z} \\ \bar{Z} \\ \vdots \\ \bar{Z} \end{pmatrix}$$

□

2.7. Medidas de la bondad de ajuste

Es deseable que los valores ajustados $\hat{Y}_i$ estén próximos a los valores observados Y_i, lo que es equivalente a decir que los residuos sean pequeños (en valor absoluto). En el caso extremo $\hat{Y}_i = Y_i$, los residuos serán todos iguales a cero, y se dice que el ajuste es perfecto. De aquí, una primera medida para valorar la bondad del ajuste es la suma de cuadrados de los residuos. Sin embargo, este criterio de bondad de ajuste (la suma de cuadros de los residuos) depende de las unidades de medida, por lo que es necesario buscar una medida alternativa adimensional.

Definición 10. *La suma de cuadrados total mide la variación o dispersión de las observaciones de la variable dependiente con respecto a su media*

$$SCT = \sum_{i=1}^{n}(Y_i - \bar{Y})^2$$

Proposición 10. *Si el modelo incluye término constante, la SCT puede expresarse como la suma de dos componentes*

$$\sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 + \sum_{i=1}^{n}\hat{u}_i^2$$

el primer componente, que se denomina suma de cuadrados explicada SCE, mide la porción de la variación total de la variable dependiente explicada por la regresión, mientras que el segundo, la suma de cuadrados de los residuos SCR, mide la variación residual o no explicada. Se tiene que

$$SCT = SCE + SCR$$

Demostración. Observando que $Y_i - \bar{Y} = Y_i - \bar{Y} + \hat{Y}_i - \hat{Y}_i$, tenemos

$$\begin{aligned}\sum_{i=1}^{n}(Y_i - \bar{Y})^2 &= \sum_{i=1}^{n}\left[(\hat{Y}_i - \bar{Y}) + (Y_i - \hat{Y}_i)\right]^2 = \sum_{i=1}^{n}\left[(\hat{Y}_i - \bar{Y})^2 + \hat{u}_i^2 + 2(\hat{Y}_i - \bar{Y})\hat{u}_i\right] \\ &= \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 + \sum_{i=1}^{n}\hat{u}_i^2\end{aligned}$$

porque $\sum_{i=1}^{n}\hat{Y}_i\hat{u}_i = 0$ e $\bar{Y}\sum_{i=1}^{n}\hat{u}_i = 0$. □

Definición 11. *Si el modelo incluye término constante, el coeficiente de determinación, R^2, es una medida de bondad de ajuste que indica la proporción de la varianza de la variable dependiente que es explicada por la regresión, y se calcula como*

$$R^2 = \frac{SCE}{SCT} = 1 - \frac{SCR}{SCT}$$

Proposición 11. *Si el modelo incluye término constante, entonces*

$$0 \leq R^2 \leq 1$$

Demostración. Si $\hat{u}_i = 0$ o $Y_i = \hat{Y}_i$ $(\forall i = 1, \ldots, n)$, entonces $SCR = 0$ y $R^2 = 1$; si $\hat{Y}_i = \bar{Y}$ o $Y_i - \bar{Y} = \hat{u}_i$ $(\forall i = 1, \ldots, n)$, entonces $SCR = SCT$ y $R^2 = 0$. Como $SCE \leq SCT$, entonces $R^2 \leq 1$; y dado que $SCR \leq SCT$, entonces $R^2 \geq 0$. □

En notación matricial, las definiciones de SCT, SCE, SCR y R^2 son

- $SCT = \mathbf{y}'\mathbf{y} - n\bar{Y}^2 = \mathbf{y}'\mathbf{M}_i\mathbf{y}$
- $SCE = \hat{\mathbf{y}}'\hat{\mathbf{y}} - n\bar{Y}^2 = \hat{\mathbf{y}}'\mathbf{M}_i\hat{\mathbf{y}}$
- $SCR = \hat{\mathbf{u}}'\hat{\mathbf{u}} = \mathbf{y}'\mathbf{M}\mathbf{y}$
- $R^2 = \dfrac{\hat{\mathbf{y}}'\mathbf{M}_i\hat{\mathbf{y}}}{\mathbf{y}'\mathbf{M}_i\mathbf{y}} = 1 - \dfrac{\mathbf{y}'\mathbf{M}\mathbf{y}}{\mathbf{y}'\mathbf{M}_i\mathbf{y}}$

Definición 12. *Los grados de libertad de un estadístico se definen como el rango de una forma cuadrática o como el número de valores que varían libremente en el cálculo de un estadístico.*

Proposición 12. *Los grados de libertad de la suma de cuadrados total son iguales a $n-1$.*

Demostración. El rango de la forma cuadrática $\mathbf{y}'\mathbf{M}_i\mathbf{y}$ es el rango de la matriz $\mathbf{M}_i$ que es igual a $n-1$. En otras palabras, para calcular la SCT tenemos n valores, pero antes hemos estimado la media, lo que nos deja $n-1$ valores variando libremente. □

Proposición 13. *Los grados de libertad de la suma de cuadrados de los residuos son iguales a $n-k$.*

Demostración. El rango de la forma cuadrática $\mathbf{y}'\mathbf{M}\mathbf{y}$ es el rango de la matriz $\mathbf{M}$ que es igual a $n-k$. Dicho de otro modo, para calcular la SCR tenemos n residuos obtenidos después de estimar k coeficientes de regresión, luego tenemos $n-k$ residuos que varían libremente. □

Proposición 14. *Los grados de libertad de la suma de cuadrados explicada son iguales a $k-1$.*

Demostración. Como la $SCE = SCT - SCR$, tenemos que

$$\hat{\mathbf{y}}'\mathbf{M}_i\hat{\mathbf{y}} = \mathbf{y}'\mathbf{M}_i\mathbf{y} - \mathbf{y}'\mathbf{M}\mathbf{y} = \mathbf{y}'\left[\mathbf{M}_i - \mathbf{M}\right]\mathbf{y}$$

Por tanto, el rango de la forma cuadrática $\hat{\mathbf{y}}'\mathbf{M}_i\hat{\mathbf{y}}$ es el rango de la matriz $\mathbf{M}_i - \mathbf{M}$ que es igual a $k-1$. De otra forma, los n valores Y_i pueden variar libremente, mientras que $n-k$ de n residuos que pueden variar libremente. Como $Y_i = \hat{Y}_i + \hat{u}_i$ se deduce que k valores ajustados pueden variar libremente. Para calcular, la SCE tenemos que calcular la media de los valores ajustados, tenemos entonces $k-1$ valores ajustados que varían libremente. □

Un inconveniente del R^2 es que nunca disminuye al aumentar el número de variables independientes en la regresión, aunque no tengan capacidad explicativa . Por tanto, este criterio de bondad de ajuste favorece a los modelos complejos que incluyen muchas variables (algunas irrelevantes) frente a los modelos simples que contienen un número reducido de variables.

Definición 13. *El coeficiente de determinación ajustado, $\bar{R}^2$, es el coeficiente de determinación corregido por los grados de libertad*

$$\bar{R}^2 = 1 - \frac{\mathbf{y}'\mathbf{M}\mathbf{y}/(n-k)}{\mathbf{y}'\mathbf{M}_i\mathbf{y}/(n-1)} = 1 - \frac{n-1}{n-k}(1-R^2)$$

Esta medida de bondad de ajuste penaliza la inclusión de variables independientes que no tengan capacidad explicativa. Así, si las variables añadidas dejan inalterado el R^2, entonces el $\bar{R}^2$ disminuirá porque el factor de penalización $(n-1)/(n-k)$ aumenta con k.

2.8. Casos especiales del modelo lineal general

Dentro de la clase general de modelos de regresión lineal, hay cuatro miembros que merecen especial atención:

1. regresión sobre una constante,
2. regresión simple,
3. regresión simple a través del origen y
4. regresión sobre dos variables.

2.8.1. Regresión sobre una constante.

Definición 14. *El modelo de regresión sobre una constante es*

$$Y_i = \beta_1 + u_i \quad i = 1, \ldots, n$$

Proposición 15. *El estimador minimocuadrático de la constante en (14) es la media muestral de la variable dependiente*

$$\hat{\beta}_1 = \bar{Y}$$

Demostración. Partiendo del modelo estimado $Y_i = \hat{\beta}_1 + u_i$ $(i = 1, \ldots, n)$ y sumando todas las observaciones obtenemos la ecuación normal

$$\sum_{i=1}^{n} Y_i = n\hat{\beta}_1$$

que podemos también obtener particularizando el sistema (2.2) para $k = 2$. Alternativamente, podemos usar la expresión general del estimador

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = (n)^{-1}\sum_{i=1}^{n} Y_i$$

en donde $\mathbf{X}$ es ahora una vector columna de unos y $\hat{\boldsymbol{\beta}}$ es el escalar $\hat{\beta}_1$. □

Proposición 16. *El coeficiente de determinación R^2 en la regresión sobre una constante es igual a cero.*

Demostración. Es claro que los residuos en la estimación minimocuadrática de (14) son los datos centrados de la variable dependiente, esto es,

$$Y_i = \hat{\beta}_1 + \hat{u}_i \Longrightarrow \hat{u}_i = Y_i - \bar{Y}$$

Por tanto, la suma de cuadrados de los residuos es igual a la suma de cuadrados total, siendo la suma de cuadrados explicada igual a cero. Logicamente, un modelo que no incluye variables explicativas no tiene capacidad explicativa. □

2.8.2. Regresión simple.

Definición 15. *El modelo de regresión lineal simple es*

$$Y_i = \beta_1 + \beta_2 X_i + u_i \quad i = 1, \ldots, n$$

en donde X_i es la variable explicativa.

Observación 7. Usamos X_i en lugar de X_{2i} porque el modelo sólo incluye una variable explicativa.

Definición 16. *La covarianza muestral entre dos variables Y y X es una medida de su asociación lineal y se define como*

$$S_{XY} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{n}$$

o, equivalentemente, como

$$S_{XY} = \frac{\sum_{i=1}^n X_i Y_i - n\bar{X}\bar{Y}}{n}$$

Observación 8. La covarianza muestral de una variable consigo misma es la varianza muestral. Así,

$$S_{XX} = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n} = \frac{\sum_{i=1}^n X_i^2 - n\bar{X}^2}{n} = S_X^2$$

Proposición 17. *En el modelo de regresión simple, los estimadores de mínimos cuadrados de β_1 y β_2 son*

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X} \qquad y \qquad \hat{\beta}_2 = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

en donde $\bar{Y}$ y $\bar{X}$ son las medias muestrales de las observaciones de Y y X.

Demostración. Para $k = 2$, las ecuaciones normales (2.2) son

$$n\hat{\beta}_1 + \hat{\beta}_2 \sum_{i=1}^n X_i = \sum_{i=1}^n Y_i$$

$$\hat{\beta}_1 \sum_{i=1}^n X_i + \hat{\beta}_2 \sum_{i=1}^n X_i^2 = \sum_{i=1}^n X_i Y_i$$

Despejando $\hat{\beta}_1$ en la primera ecuación normal

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n Y_i}{n} - \hat{\beta}_2 \frac{\sum_{i=1}^n X_i}{n} = \bar{Y} - \hat{\beta}_2 \bar{X}$$

Multiplicando la primera ecuación normal por $\sum_{i=1}^n X_i$ y la segunda por n se obtiene

$$n\hat{\beta}_1 \sum_{i=1}^n X_i + \hat{\beta}_2 \left(\sum_{i=1}^n X_i\right)^2 = \left(\sum_{i=1}^n Y_i\right)\left(\sum_{i=1}^n X_i\right)$$

$$n\hat{\beta}_1 \sum_{i=1}^n X_i + n\hat{\beta}_2 \sum_{i=1}^n X_i^2 = n \sum_{i=1}^n X_i Y_i$$

y restando la primera ecuación de la segunda tenemos

$$n\hat{\beta}_2 \left[\sum_{i=1}^n X_i^2 - \frac{1}{n}\left(\sum_{i=1}^n X_i\right)^2\right] = n\left[\sum_{i=1}^n X_i Y_i - \frac{1}{n}\left(\sum_{i=1}^n Y_i\right)\left(\sum_{i=1}^n X_i\right)\right]$$

que puede escribirse como

$$\hat{\beta}_2 \left[\sum_{i=1}^{n} X_i^2 - n\bar{X}^2 \right] = \left[\sum_{i=1}^{n} X_i Y_i - n\bar{X}\bar{Y} \right]$$

de donde

$$\hat{\beta}_2 = \frac{\sum_{i=1}^n X_i Y_i - n\bar{X}\bar{Y}}{\sum_{i=1}^n X_i^2 - n\bar{X}^2} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

□

Observación 9. *La demostración puede simplificarse usando el modelo en desviaciones*

$$y_i = \hat{\beta}_2 x_i + \hat{u}_i, \quad i = 1, \ldots, n$$

en donde $y_i = Y_i - \bar{Y}$ *y* $x_i = X_i - \bar{X}$. *De la ecuación normal*

$$\sum_{i=1}^{n} x_i y_i = \hat{\beta}_2 \sum_{i=1}^{n} x_i^2$$

obtenemos

$$\hat{\beta}_2 = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2} = \frac{S_{XY}}{S_{XX}}$$

Definición 17. *El coeficiente de correlación muestral entre dos variables* Y *y* X *es una medida adimensional de su relación lineal y se define como el cociente de su covarianza entre las respectivas desvaciones típicas*

$$r_{YX} = \frac{S_{YX}}{\sqrt{S_{YY}}\sqrt{S_{XX}}} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2}}$$

Proposición 18. *El coeficiente de determinación* R^2 *en la regresión de* Y *sobre* X *es igual al cuadrado del coeficiente de correlación muestral entre* Y *y* X

$$R^2 = r_{XY}^2$$

Demostración. El R^2 en regresión simple se define como

$$R^2 = \frac{SCE}{SCT} = \frac{\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} = \hat{\beta}_2^2 \frac{\sum_{i=1}^n (\hat{X}_i - \bar{X})^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} = \hat{\beta}_2^2 \frac{S_{XX}}{S_{YY}}$$

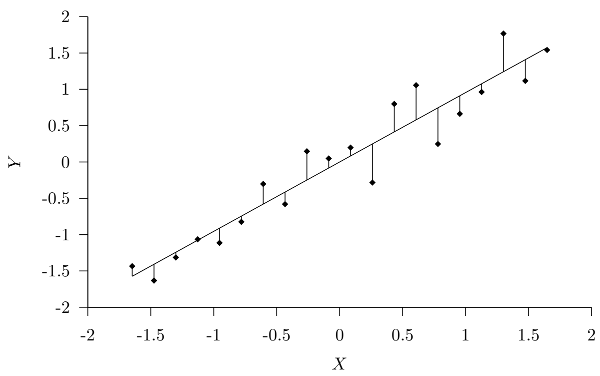
Reemplazando $\hat{\beta}_2^2$ por su expresión

$$R^2 = \frac{S_{XY}^2}{S_{XX}^2} \frac{S_{XX}}{S_{YY}} = \frac{S_{XY}^2}{S_{YY} S_{XX}} = r_{XY}^2$$

□

El análisis regresión simple se utiliza a menudo para ilustrar el ajuste de minimos cuadrados. En la figura 1 se representa una muestra de pares observaciones (Y_i, X_i), que se denomina nube de puntos, y la recta de regresión minimocuadrática. Podemos trazar o ajustar distintas rectas a través de la nube de puntos, pero sólo una de ellas minimiza los cuadrados de las distancias verticales entre cada punto y la recta. Los puntos por encima de la recta tienen asociados residuos positivos, indicando que la recta infraestima la observación; mientras que los puntos por debajo de la recta tienen asociados residuos negativos, indicando que la recta sobreestima la observación. Los puntos sobre la recta tienen asociados residuos iguales a cero. Cuando $R^2 = 0$, la recta es horizontal y los residuos son los datos centrados de Y. Cuando $R^2 = 1$, la nube de puntos se encuentra sobre la recta de regresión, siendo todos los residuos iguales a cero. La pendiente positiva de la recta de regresión indica una relación lineal positiva entre las variables, pero no puede interpretarse como una relación de causalidad.

Observación 10. *En general, la pendiente en la regresión de Y sobre X es distinta de la pendiente en la regresión de X sobre Y. Usando la figura 1, en la regresión de X sobre Y minimizamos las distancias horizontales de cada punto a la recta de regresión.*



La recta de regresión minimocuadrática se ajusta a la nube de observaciones (X_i, Y_i) minimizando los cuadrados de las distancias verticales entre cada punto y la recta. Para cada observación X_i, la recta proporciona el valor ajustado $\hat{Y}_i$. Las líneas verticales son los residuos. La recta de regresión pasa por el punto $(\bar{Y}, \bar{X})$.

Figura 1: Diagrama de dispersión y recta de regresión

2.8.3. Regresión simple a través del origen.

DEFINICIÓN 18. *El modelo de regresión simple a través del origen es*

$$Y_i = \beta X_i + u_i \quad i = 1, \ldots, n$$

en donde X_i es la variable explicativa.

Observación 11. *El nombre del modelo enfatiza que la recta de regresión pasa por el punto (0,0).*

PROPOSICIÓN 19. *En la regresión a través del origen el estimador minimocuadrático de β_2 es*

$$\hat{\beta}_2 = \frac{\sum_{i=1}^{n} X_i Y_i}{\sum_{i=1}^{n} X_i^2}$$

DEMOSTRACIÓN. El modelo estimado es

$$Y_i = \hat{\beta} X_i + \hat{u}_i \quad i = 1, \ldots, n$$

Multiplicando por X_i y sumando todas las observaciones

$$\sum_{i=1}^{n} X_i Y_i = \hat{\beta} \sum_{i=1}^{n} X_i^2 + \sum_{i=1}^{n} X_i \hat{u}_i \quad i = 1, \ldots, n$$

Como $\sum_{i=1}^{n} X_i \hat{u}_i = 0$, obtenemos que

$$\hat{\beta}_2 = \frac{\sum_{i=1}^{n} X_i Y_i}{\sum_{i=1}^{n} X_i^2}$$

Alternativamente, aplicando la fórmula general del estimador

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = (\sum_{i=1}^{n} X_i^2)^{-1} \sum_{i=1}^{n} X_i Y_i$$

en donde $\mathbf{X} = (X_1, \ldots, X_n)'$ es ahora una vector columna que contiene los datos de X y $\hat{\boldsymbol{\beta}} = (\hat{\beta})$.

□

2.8.4. Regresión lineal sobre dos variables explicativas.

Definición 19. *El modelo de regresión lineal sobre dos variables explicativas es*

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_2 X_{3i} + u_i \quad i = 1, \ldots, n$$

Proposición 20. *Los estimadores minimocuadráticos de las pendientes son*

$$\hat{\beta}_2 = \frac{\sum_{i=1}^n x_{3i}^2 \sum_{i=1}^n x_{2i}y_i - \sum_{i=1}^n x_{2i}x_{3i} \sum_{i=1}^n x_{3i}y_i}{\sum_{i=1}^n x_{2i}^2 \sum_{i=1}^n x_{3i}^2 - (\sum_{i=1}^n x_{2i}x_{3i})^2}$$
$$\hat{\beta}_3 = \frac{\sum_{i=1}^n x_{2i}^2 \sum_{i=1}^n x_{3i}y_i - \sum_{i=1}^n x_{2i}x_{3i} \sum_{i=1}^n x_{2i}y_i}{\sum_{i=1}^n x_{2i}^2 \sum_{i=1}^n x_{3i}^2 - (\sum_{i=1}^n x_{2i}x_{3i})^2}$$

Demostración. Aplicando la fórmula general del estimador para datos centrados

$$\begin{pmatrix} \hat{\beta}_2 \\ \hat{\beta}_3 \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n x_{2i}^2 & \sum_{i=1}^n x_{2i}x_{3i} \\ \sum_{i=1}^n x_{2i}x_{3i} & \sum_{i=1}^n x_{3i}^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum_{i=1}^n x_{2i}y_i \\ \sum_{i=1}^n x_{3i}y_i \end{pmatrix}$$
$$= \frac{1}{\sum_{i=1}^n x_{2i}^2 \sum_{i=1}^n x_{3i}^2 - (\sum_{i=1}^n x_{2i}x_{3i})^2} \begin{pmatrix} \sum_{i=1}^n x_{3i}^2 & -\sum_{i=1}^n x_{2i}x_{3i} \\ -\sum_{i=1}^n x_{2i}x_{3i} & \sum_{i=1}^n x_{2i}^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum_{i=1}^n x_{2i}y_i \\ \sum_{i=1}^n x_{3i}y_i \end{pmatrix}$$

□

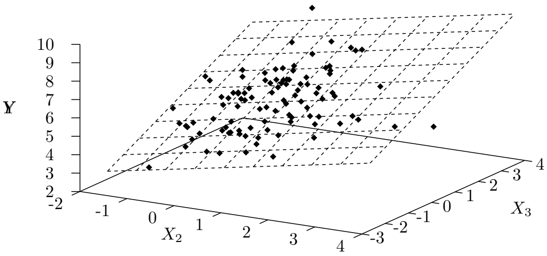
Corolario 7. *Las pendientes estimadas en la regresión sobre dos varaibles pueden expresarse en términos de las pendientes estimadas en regresiones simples*

$$\hat{\beta}_2 = \frac{\hat{\beta}_{12} - \hat{\beta}_{32}\hat{\beta}_{13}}{1 - \hat{\beta}_{23}\hat{\beta}_{32}} \quad y \quad \hat{\beta}_3 = \frac{\hat{\beta}_{13} - \hat{\beta}_{23}\hat{\beta}_{12}}{1 - \hat{\beta}_{23}\hat{\beta}_{32}}$$

en donde $\hat{\beta}_{12}$ es la pendiente estimada en la regresión de Y sobre X_2, $\hat{\beta}_{13}$ es la pendiente estimada en la regresión de Y sobre X_2, $\hat{\beta}_{23}$ es la pendiente estimada en la regresión de X_2 sobre X_3, y $\hat{\beta}_{32}$ es la pendiente estimada en la regresión de X_3 sobre X_2.

Observación 12. En general, las pendientes estimadas en la regresión sobre dos variables $\hat{\beta}_2$ y $\hat{\beta}_3$ son distintas de las pendientes estimadas en la regresión simple $\hat{\beta}_{12}$ y $\hat{\beta}_{13}$. Se cumplirá que $\hat{\beta}_2 = \hat{\beta}_{12}$ y $\hat{\beta}_3 = \hat{\beta}_{13}$ cuando $\hat{\beta}_{23} = \hat{\beta}_{32}$, es decir, cuando las variables explicativá son ortogonales.

La figura 2 representa el diagrama de dispersión tridimensional. Ahora, ajustamos un plano a la nube de puntos de manera que el cuadrado de la distancia vertical de cada punto al plano sea míninima. En el caso $k-$dimensional se dice que ajustamos un **hiperplano de regresión** a la nube de puntos de manera que el cuadrado de la distancia vertical de cada punto $(Y_i, X_{2i}, \ldots, X_{ki})$ al hiperplano sea mínima.



El plano de regresión minimocuadrático se ajusta a la nube de observaciones (Y_i, X_{2i}, X_{3i}) minimizando los cuadrados de las distancias verticales entre cada punto y el plano. Para cada par (X_{2i}, X_{3i}), el plano proporciona el valor ajustado $\hat{Y}_i$. El plano de regresión pasa por el punto $(\bar{Y}, \bar{X}_2, \bar{X}_3)$.

Figura 2: Diagrama de dispersión tridimensional y plano de regresión

2.9. Correlación simple, parcial y múltiple

El coeficiente de correlación simple es una medida de la asociación lineal entre dos variables, y toma valores entre -1 y 1. Esta medida tiene el inconveniente de estar contaminada por la influencia de otras variables. Por ejemplo, si dos variables y y x_2 no están relacionadas, pero ambas están influidas por una tercera variable x_3, el coeficiente de correlación simple no será igual a cero.

El coeficiente de correlación parcial entre y y x_2 dado x_3 es la correlación simple entre y y x_2 después de eliminar la influencia de x_3. Una de las aplicaciones del modelo de regesión es precisamente la de eliminar de una variable las influencias de terceras variables. Así, en las ecuaciones de regresión con datos centrados

$$\begin{aligned} y_i =& \hat{\beta} x_{3i} + \hat{u}_{1,3i} \\ x_{2i} =& \hat{\alpha} x_{3i} + \hat{u}_{2,3i} \end{aligned}$$

los residuos $\hat{u}_{1,3i}$ y $\hat{u}_{2,3i}$ se pueden interpretar como los componentes de y_i y x_{2i}, respectivamente, que no dependen de x_{3i}. La correlación simple entre estos residuos $\hat{u}_{1,3}$ y $\hat{u}_{2,3}$ es la correlación parcial entre y y x_2 dado x_3.

Finalmente, el coeficiente de correlación múltiple, R, es la raíz cuadrada del coeficiente de determinación.

Una vez definidos los coeficientes de correlación simple, parcial y múltiple, vamos a establecer una relación entre ellos. Consideramos, en primer lugar, la regresión lineal simple con datos centrados

$$y_i = \hat{\beta} x_{2i} + \hat{u}_{1,2i}$$

donde $\hat{u}_{1,2}$ denota los residuos en la regresión de y sobre x_2. El R^2 de esta regresión lineal simple es r_{12}^2. Y como la suma de cuadrados de los residuos $SCR = (1 - R^2)SCT$, tenemos que

$$\sum_{i=1}^{n} \hat{u}_{1,2i}^2 = (1 - r_{12}^2) \sum_{i=1}^{n} y_i^2$$

Consideramos ahora la regresión lineal simple de $\hat{u}_{1,2i}$ sobre $\hat{u}_{3,2i}$

$$\hat{u}_{1,2i} = \hat{\beta} \hat{u}_{3,2i} + \hat{u}_{1,23i}$$

esto es, la regresión del componente de y que no depende de x_2, $\hat{u}_{1,2i}$, sobre el componente de x_3 que no depende de x_2. Observamos que:

1. La SCT en esta regresión es $\sum_{i=1}^{n} \hat{u}_{1,2i}^2$,

2. El R^2 de la regresión simple es el cuadrado de la correlación simple entre $\hat{u}_{1,2}$ y $\hat{u}_{3,2}$,
3. La correlación simple entre $\hat{u}_{1,2}$ y $\hat{u}_{3,2}$ es la correlación parcial entre y y x_3 dado x_2, $r^2_{13,2}$.

Por tanto, la relación $SCR = (1 - R^2)SCT$ para esta regresión simple puede escribirse como

$$\sum_{i=1}^{n} \hat{u}^2_{1,23i} = (1 - r^2_{13,2}) \sum_{i=1}^{n} \hat{u}^2_{1,2i} = (1 - r^2_{13,2})(1 - r^2_{12}) \sum_{i=1} y_i^2$$

Análogamente, planteamos ahora la regresión simple de $\hat{u}_{1,23i}$ sobre $\hat{u}_{4,23i}$,

$$\hat{u}_{1,23i} = \hat{\beta}\hat{u}_{4,23i} + \hat{u}_{1,234i}$$

esto es, la regresión del componente de y que no depende de x_2 ni x_3, $\hat{u}_{1,23i}$, sobre el componente de x_4 que no depende de x_2 ni x_3.

Observamos que:

1. La SCT en esta regresión es $\sum_{i=1}^{n} \hat{u}^2_{1,23i}$,
2. El R^2 de la regresión simple es el cuadrado de la correlación simple entre $\hat{u}_{1,23i}$ y $\hat{u}_{4,23i}$,
3. La correlación simple entre $\hat{u}_{1,23i}$ y $\hat{u}_{4,23i}$ es la correlación parcial entre y y x_4 dados x_2 y x_3, $r^2_{14,23}$

Por tanto, la relación $SCR = (1 - R^2)SCT$ para esta ecuación de regresión puede escribirse como

$$\sum_{i=1}^{n} \hat{u}^2_{1,234i} = (1 - r^2_{14,23}) \sum_{i=1}^{n} \hat{u}^2_{1,23i} = (1 - r^2_{14,23})(1 - r^2_{13,2})(1 - r^2_{12}) \sum_{i=1} y_i^2$$

Procediendo de este modo, llegamos a

$$\sum_{i=1}^{n} \hat{u}^2_{1,2\ldots ki} = (1-r^2_{1k,2\ldots(k-1)}) \sum_{i=1}^{n} \hat{u}^2_{1,2\ldots(k-1)i} = (1-r^2_{1k,2\ldots(k-1)}) \ldots (1-r^2_{12,23})(1-r^2_{13,2})(1-r^2_{12}) \sum_{i=1} y_i^2$$

y dividiendo ambos lados de la ecuación por la SCT, tenemos

$$\frac{\sum_{i=1}^{n} \hat{u}^2_{1,2\ldots ki}}{\sum_{i=1} y_i^2} = 1 - R^2_{1,2\ldots k} = (1 - r^2_{12})(1 - r^2_{13,2})(1 - r^2_{14,23}) \ldots (1 - r^2_{1k,2\ldots(k-1)})$$

2.10. Ejemplo

El cuadro 1.a contiene las notas de 10 alumnos que se examinaron de Econometría en el curso 2006-2007. Al comienzo del curso 2007-2008 el profesor preguntó a los nuevos estudiantes qué causas podrían explicar las diferencias en esas notas. Los alumnos apuntaron rápidamente tres: las horas semanales que dedican al estudio, la asistencia regular a las clases, y la asistencia a clases particulares en una academia. Además, pensaban que el número de horas de estudio influye positivamente en la nota, como también lo hacen las otras dos variables. La asistencia regular a clase se representa por un variable dicotómica que toma el valor 1 si el alumno asiste regularmente a clase y el valor 0 en caso contrario. Análogamente, se mide la asistencia a clases particulares.

Para medir la posible influencia de estas tres variables explicativas sobre la variable dependiente formulamos el modelo de regresión lineal multivariante

$$N_i = \beta_1 + \beta_2 H_i + \beta_3 C_i + \beta_4 A_i + u_i, \quad i = 1, \ldots, 10$$

Alumno	Nota	Horas	Clase	Academia	Nota	Horas	Clase	Academia	$\hat{u}_i$
1	5	1	1	0	5	1	1	0	1.10811
2	9	5	1	0	9	5	1	0	-0.945946
3	7.5	3	1	0	7.5	3	1	0	0.581081
4	4.5	1	1	0	4.5	1	1	0	0.608108
5	4	1	1	1	4	1	1	1	-2.4869e-014
6	8	4	0	1	8	4	0	1	0.986486
7	0	0	0	1	0	0	0	1	-0.959459
8	4	2	1	0	4	2	1	0	-1.40541
9	8.5	5	0	1	8.5	5	0	1	-0.027027
10	10	5	1	0	10	5	1	0	0.0540541

(a) Datos observados (b) Datos centrados (c) Residuos

Cuadro 1: Calificaciones y otras características individuales

en donde denotamos por sus iniciales las variables del cuadro 1.a. Siguiendo la regla práctica para especificar las ecuaciones normales obtenemos

$$\begin{pmatrix} 10 & \sum_{i=1}^{10} H_i & \sum_{i=1}^{10} C_i & \sum_{i=1}^{10} A_i \\ \sum_{i=1}^{10} H_i & \sum_{i=1}^{10} H_i^2 & \sum_{i=1}^{10} H_iC_i & \sum_{i=1}^{10} H_iA_i \\ \sum_{i=1}^{10} C_i & \sum_{i=1}^{10} C_iH_i & \sum_{i=1}^{10} C_i^2 & \sum_{i=1}^{10} C_iA_i \\ \sum_{i=1}^{10} A_i & \sum_{i=1}^{10} A_iH_i & \sum_{i=1}^{10} A_iC_i & \sum_{i=1}^{10} A_i^2 \end{pmatrix} \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \hat{\beta}_4 \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{10} N_i \\ \sum_{i=1}^{10} H_iN_i \\ \sum_{i=1}^{10} C_iN_i \\ \sum_{i=1}^{10} A_iN_i \end{pmatrix}$$

y calculando todos estos momentos muestrales con los datos del cuadro 1.1, podemos encontrar las estimaciones minimocuadráticas de los parámetros

$$\begin{pmatrix} 10 & 27 & 7 & 4 \\ 27 & 107 & 18 & 10 \\ 7 & 18 & 7 & 1 \\ 4 & 10 & 1 & 4 \end{pmatrix} \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \hat{\beta}_4 \end{pmatrix} = \begin{pmatrix} 60{,}5 \\ 213{,}5 \\ 44 \\ 20{,}5 \end{pmatrix} \Rightarrow \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \hat{\beta}_4 \end{pmatrix} = \begin{pmatrix} 0{,}851351 \\ 1{,}51351 \\ 1{,}52703 \\ 0{,}108108 \end{pmatrix}$$

Estas estimaciones nos dicen lo siguiente:

$\hat{\beta}_1 = 0{,}85$: es la nota de un alumno que no estudie, que no asista a clase y que no reciba clases particulares,

$\hat{\beta}_2 = 1{,}51$: es el incremento en la nota por cada hora de estudio,

$\hat{\beta}_3 = 1{,}52$: es el incremento en la nota por asistir regularmente a clase,

$\hat{\beta}_4 = 0{,}11$: es el incremento en la nota por ir a una academia.

El modelo estimado puede usarse para predecir la nota que obtendrá un alumno que estudia 3 horas a la semana, asiste regularmente a clase y no recibe clases particulares

$$\hat{N}_i = 0{,}851351 + 1{,}51351 \times 3 + 1{,}52703 \times 1 + 0{,}108108 \times 0 = 6{,}92$$

o para estimar cuántas horas debe estudiar este alumno si quiere obtener una nota igual a 9

$$\hat{H}_i = (9 - 0{,}851351 - 1{,}52703 \times 1 - 0{,}108108 \times 0)/1{,}51351 = 4{,}38$$

El modelo puede estimarse usando los datos centrados del cuadro 1.b. Ahora el sistema de ecuaciones normales es

$$\begin{pmatrix} \sum_{i=1}^{10} h_i^2 & \sum_{i=1}^{10} h_ic_i & \sum_{i=1}^{10} h_ia_i \\ \sum_{i=1}^{10} c_ih_i & \sum_{i=1}^{10} c_i^2 & \sum_{i=1}^{10} c_ia_i \\ \sum_{i=1}^{10} a_ih_i & \sum_{i=1}^{10} a_ic_i & \sum_{i=1}^{10} a_i^2 \end{pmatrix} \begin{pmatrix} \hat{\beta}_2 \\ \hat{\beta}_3 \\ \hat{\beta}_4 \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{10} h_in_i \\ \sum_{i=1}^{10} c_in_i \\ \sum_{i=1}^{10} a_in_i \end{pmatrix}$$

en donde las letras minúsculas son las iniciales de las variables en desviaciones. Calculando estos momentos centrados obtenemos

$$\begin{pmatrix} 34{,}1 & -0{,}9 & -0{,}8 \\ -0{,}9 & 2{,}1 & -1{,}8 \\ -0{,}8 & -1{,}8 & 2{,}4 \end{pmatrix} \begin{pmatrix} \hat{\beta}_2 \\ \hat{\beta}_3 \\ \hat{\beta}_4 \end{pmatrix} = \begin{pmatrix} 50{,}15 \\ 1{,}65 \\ -3{,}7 \end{pmatrix} \Rightarrow \begin{pmatrix} \hat{\beta}_2 \\ \hat{\beta}_3 \\ \hat{\beta}_4 \end{pmatrix} = \begin{pmatrix} 1{,}51351 \\ 1{,}52703 \\ 0{,}108108 \end{pmatrix}$$

La estimación del término constante se obtiene de la ecuación que relaciona las medias muestrales

$$\hat{\beta}_1 = \bar{N} - \hat{\beta}_2\bar{H} - \hat{\beta}_2\bar{C} - \hat{\beta}_2\bar{A}$$

que conduce a

$$\hat{\beta}_1 = 6{,}05 - 1{,}51351 \times 2{,}7 - 1{,}52703 \times 0{,}7 - 0{,}108108 \times 0{,}4 = 0{,}851359$$

Vemos que usando los datos centrados obtenemos los mismos resultados, pero invertimos una matriz de menor orden. A partir de estas estimaciones podemos calcular los residuos que se presentan en el cuadro 1.c

$$\hat{u}_i = N_i - 0{,}851351 - 1{,}51351H_i - 1{,}52703C_i - 0{,}108108A_i$$

La bondad del ajuste medida por el coeficiente de determinación

$$R^2 = 1 - \frac{\sum_{i=1}^{10} \hat{u}_i^2}{\sum_{i=1}^{10} n_i^2} = 1 - \frac{6{,}7027}{84{,}725} = 0{,}920889$$

indica que la regresión explica el 92,1 % de la varianza de la variable Nota.

El análisis de regresión anterior puede realizarse por etapas incluyendo secuencialmente las variables explicativas. Usando los datos centrados del cuadro 1.b, estimamos primero la regresión simple

$$y_i = \hat{\alpha}x_{2i} + \hat{u}_{1,2i}$$

en donde $y_i = n_i$ y $x_{2i} = h_i$. De aquí obtenemos

$$\hat{\alpha} = \frac{\sum_{i=1}^n x_{2i}y_i}{\sum_{i=1}^n x_{2i}^2} = \frac{50{,}15}{34{,}1} = 1{,}47067 \quad \text{y} \quad r_{12}^2 = 1 - \frac{\sum_{i=1}^n \hat{u}_{1,2i}^2}{\sum_{i=1}^n y_i^2} = 1 - \frac{10{,}97067}{84{,}725} = 0{,}87051$$

y los residuos $\hat{u}_{1,2i}$ mostrados en el cuadro 2. Después, estimamos la regresión simple

$$x_{3i} = \hat{\alpha}x_{2i} + \hat{u}_{3,2i}$$

en donde $x_{3i} = c_i$, y obtenemos

$$\hat{\alpha} = \frac{\sum_{i=1}^n x_{2i}x_{3i}}{\sum_{i=1}^n x_{2i}^2} = \frac{-0{,}9}{34{,}1} = -0{,}026393$$

y los residuos $\hat{u}_{3,2i}$ mostrados en el cuadro 2. Ahora, podemos estimar la regresión simple

$$\hat{u}_{1,2i} = \hat{\alpha}\hat{u}_{3,2i} + \hat{u}_{1,23i}$$

obteniendo

$$\hat{\alpha} = \frac{\sum_{i=1}^n \hat{u}_{3,2i}\hat{u}_{1,2i}}{\sum_{i=1}^n \hat{u}_{3,2i}^2} = \frac{2{,}97361}{2{,}07625} = 1{,}4322 \quad \text{y} \quad r_{13,2}^2 = 1 - \frac{\sum_{i=1}^n \hat{u}_{1,23i}^2}{\sum_{i=1}^n \hat{u}_{1,2i}^2} = 1 - \frac{6{,}71186}{10{,}970674} = 0{,}388199$$

Finalmente, estimamos la regresión múltiple

$$x_{4i} = \hat{\alpha}_2x_{2i} + \hat{\alpha}_3x_{3i} + \hat{u}_{4,23i}$$

Alumno	$\hat{u}_{1,2}$	$\hat{u}_{3,2}$	$\hat{u}_{1,23}$	$\hat{u}_{4,23}$	$\hat{u}_{1,234}$
1	1.45015	0.255132	1.08475	-0.216102	1.10811
2	-0.432551	0.360704	-0.949153	-0.029661	-0.945946
3	1.0088	0.307918	0.567797	-0.122881	0.581081
4	0.950147	0.255132	0.584746	-0.216102	0.608108
5	0.450147	0.255132	0.0847458	0.783898	8.88178e-016
6	0.0381232	-0.665689	0.991525	0.0466102	0.986486
7	-2.07918	-0.771261	-0.974576	-0.139831	-0.959459
8	-1.02053	0.281525	-1.42373	-0.169492	-1.40541
9	-0.932551	-0.639296	-0.0169492	0.0932203	-0.027027
10	0.567449	0.360704	0.0508475	-0.029661	0.0540541

Cuadro 2: Residuos de diferentes regresiones

en donde $x_{4i} = a_i$, y obtenemos

$$\begin{pmatrix} \hat{\alpha}_2 \\ \hat{\alpha}_3 \end{pmatrix} = \begin{pmatrix} 34{,}1 & -0{,}9 \\ -0{,}9 & 2{,}1 \end{pmatrix}^{-1} \begin{pmatrix} -0{,}8 \\ -1{,}8 \end{pmatrix} = \frac{1}{70{,}8} \begin{pmatrix} 2{,}1 & 0{,}9 \\ 0{,}9 & 34{,}1 \end{pmatrix} \begin{pmatrix} -0{,}8 \\ -1{,}8 \end{pmatrix} = \begin{pmatrix} -0{,}0466102 \\ -0{,}877119 \end{pmatrix}$$

y los residuos $\hat{u}_{4,23i}$ mostrados en el cuadro 2. Ahora, podemos estimar la regresión simple

$$\hat{u}_{1,23i} = \hat{\alpha}\hat{u}_{4,23i} + \hat{u}_{1,234i}$$

obteniendo que

$$\hat{\alpha} = \frac{\sum_{i=1}^n \hat{u}_{3,2i}\hat{u}_{1,2i}}{\sum_{i=1}^n \hat{u}_{3,2i}^2} = \frac{0{,}0847458}{0{,}783898} = 0{,}108108 \quad \text{y} \quad r_{14,23}^2 = 1 - \frac{\sum_{i=1}^n \hat{u}_{1,234i}^2}{\sum_{i=1}^n \hat{u}_{1,23i}^2} = 1 - \frac{6{,}7027}{6{,}71186} = 0{,}001365$$

y los residuos $\hat{u}_{1,234i}$ mostrados en el cuadro 2. Vemos que la estimación $\hat{\alpha} = 0{,}108108$ y los residuos $\hat{u}_{1,234i}$ de esta regresión simple coinciden con la estimación de $\hat{\beta}_4$ y los residuos $\hat{u}_i$ obtenidos en la regresión múltiple.

Podemos comprobar que

$$\begin{aligned} 1 - R_{1,234}^2 =& (1 - r_{12}^2)(1 - r_{13,2}^2)(1 - r_{14,23}^2) \\ 1 - 0{,}92088873 =& (1 - 0{,}870514)(1 - 0{,}388199)(1 - 0{,}001365) \\ 0{,}0791113 =& 0{,}129486 \times 0{,}611801 \times 0{,}998635 \end{aligned}$$

Resumen

1. El método de mínimos cuadrados tiene la propiedad numérica fundamental de generar residuos ortogonales a los regresores: $\mathbf{X}'\hat{\mathbf{u}} = \mathbf{0}$.
2. Las ecuaciones normales $\mathbf{X}'\mathbf{y} = \mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}}$ pueden obtenerse premultiplicando el modelo estimado $\mathbf{y} = \mathbf{X}\hat{\boldsymbol{\beta}} + \hat{\mathbf{u}}$ por la matriz $\mathbf{X}'$.
3. El estimador de mínimos cuadrados ordinarios (MCO) es $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$.
4. Utilizando datos centrados reducimos en una unidad la dimensión de la matriz $(\mathbf{X}'\mathbf{X})^{-1}$.
5. La regresión con datos tipificados permite apreciar la importancia relativa de cada variable explicativa.
6. El coeficiente de determinación R^2 mide la proporción de la varianza del regresando explicada por los regresores.

7. En la regresión simple de Y sobre X, el R^2 es el cuadrado del coeficiente de correlación muestral entre Y y X.
8. Los coeficientes de regresión estimados por MCO son coeficientes de regresión parcial, miden la relación entre el regresando y el regresor después de extraer la influencia de los otros regresores.

Palabras clave

Observaciones	Suma de cuadrados total
Valores ajustados	Suma de cuadrados explicada
Residuos	Suma de cuadrados residual
Ecuaciones normales	Coeficiente de determinación
Estimaciones minimocuadráticas	Coeficientes de regresión parcial

Ejercicios

1. Considere tres variables Y, X y Z. Escriba la ecuación de regresión múltiple de Z sobre un término constante, Y y X.
2. Sea $Y_t\,(t = 1, \ldots, 10)$ una serie temporal descrita por ecuación de regresión múltiple

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + u_t, \quad t = 3, \ldots, 10$$

Escriba el modelo de regresión en forma matricial, indicando los elementos del vector $\mathbf{y}$ y de la matriz $\mathbf{X}$.
3. El cuadro 3 contiene las series temporales de Consumo Nacional (C_t) y Renta Nacional Disponible (Y_t) en España para el periodo 1995-2005 a precios corrientes. Usando estos datos, obtenga las ecuaciones normales y las estimaciones de mínimos cuadrados para el modelo de regresión simple $C_t = \beta_1 + \beta_2 Y_t + u_t$.

Año	CN	RD
1995	349.268	388.029
1996	368.474	408.24
1997	388.373	433.777
1998	414.158	465.222
1999	444.982	498.647
2000	484.359	538.594
2001	518.484	574.786
2002	550.49	614.7
2003	586.594	656.77
2004	635.993	699.271
2005	686.699	748.357

Cuadro 3: Consuno y Renta a precios corrientes (10^9 euros)

4. Para la regresión lineal simple estimada en el ejercicio anterior, calcule las sumas de cuadrados total, explicada y residual, así como el coeficiente de determinación. Compruebe que el R^2 es el cuadrado del coeficiente de correlación simple entre C_t e Y_t.

5. Sea $\hat{\mathbf{u}}$ el vector de residuos en la regresión de $\mathbf{y}$ sobre $\mathbf{X}$. Demuestre que el estimador de mínimos cuadrados en la regresión de $\hat{\mathbf{u}}$ sobre $\mathbf{X}$ es un vector nulo.
6. Sea $\mathbf{M}$ la matriz de proyección generadora de residuos de orden $n \times n$ y sea $\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_m)$ una matriz $n \times m$. ¿Cómo se intepreta el producto $\mathbf{MY}$?
7. Considere los modelos de regresión

$$\begin{aligned}\mathbf{y} =& \mathbf{X}_r\boldsymbol{\beta}_r + \mathbf{e}\\ \mathbf{y} =& \mathbf{X}_r\boldsymbol{\beta}_r + \mathbf{X}_s\boldsymbol{\beta} + \mathbf{u}\end{aligned}$$

   Escriba la expresión del estimador de mínimos cuadrados de $\boldsymbol{\beta}_r$ en cada modelo. ¿En qué caso especial ambos estimadores serán iguales?
8. Demuestre que, para los modelos del ejercicio anterior, $\hat{\mathbf{u}}'\hat{\mathbf{u}} \leq \hat{\mathbf{e}}'\hat{\mathbf{e}}$.
9. Utilizando los datos del cuadro 3 estime por mínimos cuadrados las ecuaciones de regresión

$$\begin{aligned}C_t =& \alpha_1 + \beta_1 Y_t + u_{1t}\\ C_t =& \alpha_2 + \beta_2 C_{t-1} + u_{2t}\\ C_{t-1} =& \alpha_3 + \beta_3 Y_t + u_{3t}\\ Y_t =& \alpha_4 + \beta_4 C_{t-1} + u_{4t}\end{aligned}$$

   Use estas estimaciones para calcular las estimaciones de γ_2 y γ_3 en

$$C_t = \gamma_1 + \gamma_2 Y_t + \gamma_3 C_{t-1} + u_t$$

10. Dos investigadores han utilizado muestras diferentes sobre las mismas variables para estimar por mínimos cuadrados la ecuación de regresión simple

$$y_i = \beta_1 + \beta_2 X_i + u_i$$

    Deciden unir esfuerzos y se plantean dos estrategias:
    a) calcular la media aritmética de sus estimaciones,
    b) reestimar los parámetros utilizando todos los datos.
    ¿Qué procedimiento produciría una menor suma de cuadrados de los residuos?
11. Demuestre que el coeficiente de determinación es invariante a transformaciones lineales de las variables.