Modelos Lineales

Regresión Lineal Simple

Fernando Massa; Bruno Bellagamba

Martes 12 de marzo 2023

FCEA Facultad de Ciencias Económicas y de Administración
IESTA Instituto de Estadística
Universidad de la República Uruguay

Temario

Objetivos

- Obtendremos el estimador MCO en su forma matricial.
- Indagaremos sobre su $\mathbb{E}$ y $\mathbb{V}ar$.
- Exploraremos las *sumas de cuadrados* y a partir de estas definiremos el coeficiente de determinación.
- Nos iremos familiarizando con el entorno de trabajo del R.
- Haremos algún ejercicio del Práctico 1.

La clase pasada obtuvimos el estimador $\hat{\beta}_{MCO}$ al resolver las ecuaciones normales. Hoy retomaremos este problema desde una punto de vista matricial.

La clase pasada obtuvimos el estimador $\hat{\beta}_{MCO}$ al resolver las ecuaciones normales. Hoy retomaremos este problema desde una punto de vista matricial.

En primer lugar, debemos plantear el modelo de RLS desde un punto de vista matricial. Recordémoslo que en su forma escalar:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \qquad \forall i = 1, \ldots, n$$

MCO

La clase pasada obtuvimos el estimador $\hat{\beta}_{MCO}$ al resolver las ecuaciones normales. Hoy retomaremos este problema desde una punto de vista matricial.

En primer lugar, debemos plantear el modelo de RLS desde un punto de vista matricial. Recordémoslo que en su forma escalar:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \qquad \forall i = 1, \ldots, n$$

Debido a que contamos con n ecuaciones como la anterior (una para cada observación), podemos expresar el modelo de la siguiente manera:

$$Y = X\beta + \varepsilon$$

Donde Y y ε son vectores de $n \times 1$, X es una matriz de $n \times 2$ y β es un vector de 2×1

La clase pasada obtuvimos el estimador $\hat{\beta}_{MCO}$ al resolver las ecuaciones normales. Hoy retomaremos este problema desde una punto de vista matricial.

En primer lugar, debemos plantear el modelo de RLS desde un punto de vista matricial. Recordémoslo que en su forma escalar:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \qquad \forall i = 1, \ldots, n$$

Debido a que contamos con n ecuaciones como la anterior (una para cada observación), podemos expresar el modelo de la siguiente manera:

$$Y = X\beta + \varepsilon$$

Donde Y y ε son vectores de $n \times 1$, X es una matriz de $n \times 2$ y β es un vector de 2×1

Veámoslo un poco más detalladamente

MCO

A partir de una conjunto de n observaciones, postulamos la relación lineal para cada una de ellas de la siguiente manera:

$$\begin{aligned} y_1 &= \beta_0 + \beta_1 x_1 + \varepsilon_1 \\ y_2 &= \beta_0 + \beta_1 x_2 + \varepsilon_2 \\ \vdots &= \vdots \\ y_n &= \beta_0 + \beta_1 x_n + \varepsilon_n \end{aligned}$$

Lo cual es equivalente a:

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

Y así expresamos el modelo lineal de manera más compacta como:

$$Y = X\beta + \varepsilon$$

MCO

La clase pasada, vimos que el problema de estimar el vector de coeficientes β comenzaba con la minimización de la función:

$$\min_{\beta_0,\beta_1} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2 = \sum_{i=1}^{n} \varepsilon_i^2$$

MCO

La clase pasada, vimos que el problema de estimar el vector de coeficientes β comenzaba con la minimización de la función:

$$\min_{\beta_0,\beta_1} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2 = \sum_{i=1}^{n} \varepsilon_i^2$$

Matricialmente, esto equivale a:

$$\min_{\beta} (Y - X\beta)^{'} (Y - X\beta) = \varepsilon^{'} \varepsilon$$

MCO

La clase pasada, vimos que el problema de estimar el vector de coeficientes β comenzaba con la minimización de la función:

$$\min_{\beta_0,\beta_1} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2 = \sum_{i=1}^{n} \varepsilon_i^2$$

Matricialmente, esto equivale a:

$$\min_{\beta} (Y - X\beta)^{'} (Y - X\beta) = \varepsilon^{'} \varepsilon$$

Para obtener el mínimo, derivamos (vectorialmente) según β e igualamos al vector nulo

$$\frac{\partial}{\partial \beta} (Y - X\beta)^{'} (Y - X\beta) = 0$$

MCO

Antes de aplicar la propiedad la distributiva, recordemos que:

Propiedades de matrices

- $(AB)^{'} = B^{'}A^{'}$
- $\frac{\partial}{\partial x}a^{'}x = \frac{\partial}{\partial x}x^{'}a = a$
- $\frac{\partial}{\partial x}x^{'}Ax = 2Ax$

Antes de aplicar la propiedad la distributiva, recordemos que:

Propiedades de matrices

- $(AB)^{'} = B^{'}A^{'}$
- $\frac{\partial}{\partial x} a^{'}x = \frac{\partial}{\partial x} x^{'}a = a$
- $\frac{\partial}{\partial x} x^{'}Ax = 2Ax$

Retomando la última ecuación de la diapositiva anterior:

$$\frac{\partial}{\partial \beta}(Y - X\beta)^{'}(Y - X\beta) = 0$$
$$\frac{\partial}{\partial \beta}\left(Y^{'}Y - Y^{'}X\beta - \beta^{'}X^{'}Y + \beta^{'}X^{'}X\beta\right) = 0$$

Luego de derivar cada sumando por separado se obtiene:

$$-X^{'}Y - X^{'}Y + 2X^{'}X\beta = 0$$

MCO

Retomando el resultado de la diapositiva anterior

$$-2X^{'}Y + 2X^{'}X\beta = 0$$

MCO

Retomando el resultado de la diapositiva anterior

$$-2X^{'}Y + 2X^{'}X\beta = 0$$

Al despejar, obtenemos que: $$X^{'}X\beta = X^{'}Y$$

MCO

Retomando el resultado de la diapositiva anterior

$$-2X^{'}Y + 2X^{'}X\beta = 0$$

Al despejar, obtenemos que: $$X^{'}X\beta = X^{'}Y$$

El mismo resultado puede obtenerse sin la cuarta propiedad, si te tiene en cuanta que los 4 términos resultantes de aplicar la propiedad distributiva son escalares.

$$Y^{'}Y - Y^{'}X\beta - \beta^{'}X^{'}Y + \beta^{'}X^{'}X\beta$$

Retomando el resultado de la diapositiva anterior

$$-2X^{'}Y+2X^{'}X\beta=0$$

Al despejar, obtenemos que: $$X^{'}X\beta=X^{'}Y$$

El mismo resultado puede obtenerse sin la cuarta propiedad, si te tiene en cuanta que los 4 términos resultantes de aplicar la propiedad distributiva son escalares.

$$Y^{'}Y-Y^{'}X\beta-\beta^{'}X^{'}Y+\beta^{'}X^{'}X\beta$$

Como el transpuesto de un escalar es el propio escalar, entonces $\beta^{'}X^{'}Y=Y^{'}X\beta$. De lo cual resulta que la expresión anterior se puede escribir como:

$$Y^{'}Y-2X^{'}Y\beta+\beta^{'}X^{'}X\beta$$

Facilitando la derivación.

No obstante, para poder realizar el último paso y obtener el estimador se hace necesario un supuesto.

RANGO DE X

La matriz X debe ser de rango completo. Es decir, todas sus columnas deben formar un conjunto linealmente independiente (Asumiento $k << n$). En el caso del modelo de RLS, el rango de X debe ser 2.

No obstante, para poder realizar el último paso y obtener el estimador se hace necesario un supuesto.

Rango de X

La matriz X debe ser de rango completo. Es decir, todas sus columnas deben formar un conjunto linealmente independiente (Asumiento $k << n$). En el caso del modelo de RLS, el rango de X debe ser 2.

Este supuesto nos garantiza que $X^{'}X$ posee inversa. Entonces el estimador es:

$$\hat{\beta}_{MCO} = \left(X^{'}X\right)^{-1} X^{'}Y$$

Momentos del estimador $\hat{\beta}_{MCO}$

Para analizar las propiedades del estimador, es conveniente retomar el modelo en su expresión matricial.

$$Y = X\beta + \varepsilon$$

Momentos del estimador $\hat{\beta}_{MCO}$

Para analizar las propiedades del estimador, es conveniente retomar el modelo en su expresión matricial.

$$Y = X\beta + \varepsilon$$

Anteriormente hicimos los siguientes supuestos sobre cada uno de los componentes del vector de errores:

- $E(\varepsilon_i) = 0 \qquad \forall i = 1, \dots, n$
- $Var(\varepsilon_i) = \sigma^2 \qquad \forall i = 1, \dots, n$
- $Cov(\varepsilon_i, \varepsilon_j) = 0 \qquad \forall i \neq j$

Matricialmente esto equivale a decir que:

- $E(\varepsilon) = \mathbf{0}$
- $Cov(\varepsilon) = \sigma^2 I_n$

Momentos del estimador $\hat{\beta}_{MCO}$

Entonces, los momentos del vector Y son:

- $E(Y) = X\beta$
- $Cov(Y) = \sigma^2 I_n$

Entonces, los momentos del vector Y son:

- $E(Y) = X\beta$
- $Cov(Y) = \sigma^2 I_n$

Teniendo esto en cuenta, podemos calcular de manera sencilla $\mathbb{E}(\hat{\beta}_{MCO})$

$$\begin{aligned}
\mathbb{E}(\hat{\beta}_{MCO}) &= \mathbb{E}\left[\left(X'X\right)^{-1} X'Y\right] \\
&= \left(X'X\right)^{-1} X'\mathbb{E}(Y) \\
&= \left(X'X\right)^{-1} X'X\beta \\
&= \beta
\end{aligned}$$

Por lo cual, nuestro estimador es **insesgado**.

Momentos del estimador $\hat{\beta}_{MCO}$

Para calcular la varianza del estimador, es necesario recordar que:

Varianza de una transformación lineal

Sean $X \in \mathbb{R}^n$ una variable aleatoria con matriz de covarianzas Σ y sean $A \in \mathbb{R}^{p \times n}$ y $b \in \mathbb{R}^p$ una matriz y un vector de coeficientes fijos, entonces:

$$\mathbb{V}ar(AX + b) = A\Sigma A^{'}$$

Momentos del estimador $\hat{\beta}_{MCO}$

Para calcular la varianza del estimador, es necesario recordar que:

Varianza de una transformación lineal

Sean $X \in \mathbb{R}^n$ una variable aleatoria con matriz de covarianzas Σ y sean $A \in \mathbb{R}^{p \times n}$ y $b \in \mathbb{R}^p$ una matriz y un vector de coeficientes fijos, entonces:

$$\mathbb{V}ar(AX + b) = A\Sigma A^{'}$$

Teniendo esto en cuenta, podemos calcular $\mathbb{V}ar(\hat{\beta}_{MCO})$

$$\begin{aligned}
\mathbb{V}ar(\hat{\beta}_{MCO}) &= \mathbb{V}ar\left[\left(X^{'}X\right)^{-1}X^{'}Y\right] \\
&= \left(X^{'}X\right)^{-1}X^{'}\mathbb{V}ar(Y)X\left(X^{'}X\right)^{-1} \\
&= \left(X^{'}X\right)^{-1}X^{'}\sigma^2 I_n X\left(X^{'}X\right)^{-1} \\
&= \sigma^2\left(X^{'}X\right)^{-1}X^{'}X\left(X^{'}X\right)^{-1} \\
&= \sigma^2\left(X^{'}X\right)^{-1}
\end{aligned}$$

Momentos del estimador $\hat{\beta}_{MCO}$

Recapitulando:

- $\mathbb{E}(\hat{\beta}_{MCO}) = \beta$
- $\mathbb{V}ar(\hat{\beta}_{MCO}) = \sigma^2 \left(X'X\right)^{-1}$

Recapitulando:

- $\mathbb{E}(\hat{\beta}_{MCO}) = \beta$
- $\mathbb{V}ar(\hat{\beta}_{MCO}) = \sigma^2 \left(X^{'} X\right)^{-1}$

No obstante, debido a que en la expresión de la varianzade $\hat{\beta}_{MCO}$ aparece el parámetro desconocido σ^2, *estimaremos* la varianza del estimador mediante:

$$\widehat{\mathbb{V}ar(\hat{\beta}_{MCO})} = \hat{\sigma}^2 \left(X^{'} X\right)^{-1}$$

Recapitulando:

- $\mathbb{E}(\hat{\beta}_{MCO}) = \beta$
- $\mathbb{V}ar(\hat{\beta}_{MCO}) = \sigma^2 \left(X'X\right)^{-1}$

No obstante, debido a que en la expresión de la varianzade $\hat{\beta}_{MCO}$ aparece el parámetro desconocido σ^2, *estimaremos* la varianza del estimador mediante:

$$\widehat{\mathbb{V}ar(\hat{\beta}_{MCO})} = \hat{\sigma}^2 \left(X'X\right)^{-1}$$

Siendo: $$\hat{\sigma}^2 = \frac{1}{n - rg(X)} \left(Y - X\hat{\beta}_{MCO}\right)' \left(Y - X\hat{\beta}_{MCO}\right)$$

Más adelante veremos la razón por la cuál restamos el $rg(X)$ en el denominador de este estadístico.

Momentos del estimador $\hat{\beta}_{MCO}$

Si bien profundizaremos sobre esto más adelante, los elementos de la diagonal de esta matriz nos servirán para realizar pruebas de hipótesis del estilo:

$$\begin{aligned} H_0) &\quad \beta_j = \beta^{(0)} \\ H_1) &\quad \beta_j \neq \beta^{(0)} \end{aligned}$$

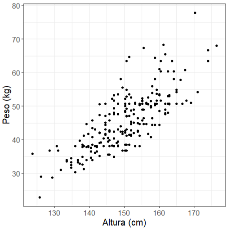
Mediante el estadístico de Wald:

$$T = \frac{\hat{\beta}_j - \beta^{(0)}}{s_{\hat{\beta}_j}}$$

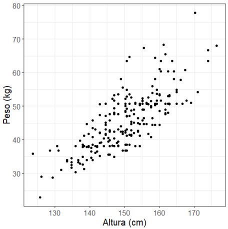
Siendo $s_{\hat{\beta}_j}$ el *j-ésimo* elemento de la diagonal de nuestra estimación de la matriz de varianzas y covarianzas de $\hat{\beta}_{MCO}$.

Bajo $H_0)$ cierta, este estadístico se distribuye $t_{n-rg(X)}$.

Uno de los ejemplos mencionados en la clase anterior suponía modelar la (posible) relación lineal entre el peso y la altura en un conjunto de 263 adolescentes.

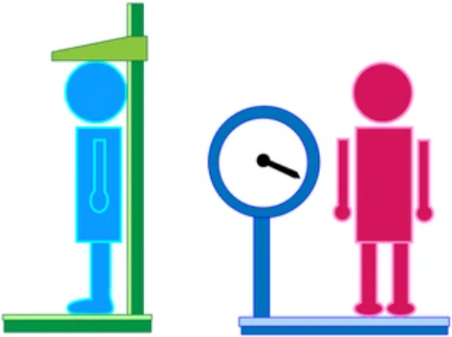


Uno de los ejemplos mencionados en la clase anterior suponía modelar la (posible) relación lineal entre el peso y la altura en un conjunto de 263 adolescentes.



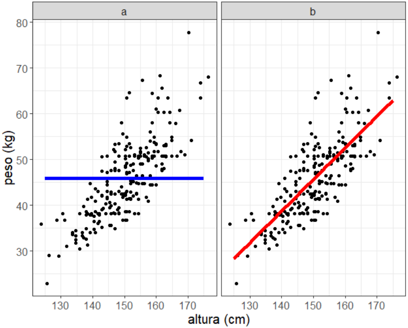
Si quisiéramos modelar el peso en función de la altura, el modelo de RLS sería:

$$peso_i = \beta_0 + \beta_1 altura_i + \varepsilon_i$$



Vayamos al R y ejecutemos el script *peso_altura.r*

A la hora de determinar la calidad del ajuste del modelo RLS, surge la interrogante de comparar el aporte del coeficiente asociado a la pendiente con respecto de no utilizarlo. En el ejemplo de la altura, la diferencia estaría entre estas dos situaciones.



Descomposición de la variabilidad

La idea de la figura es que podemos descomponer la distancia de un punto a la media como la distancia del punto a la recta + la distancia de la recta a la media.

La idea de la figura es que podemos descomponer la distancia de un punto a la media como la distancia del punto a la recta + la distancia de la recta a la media.

Las distancias de todos los puntos a la recta azul (la media) es:

$$\sum_i (y_i - \bar{y})^2$$

La cual es llamada **Suma de Cuadrados Total** (SCT).

La idea de la figura es que podemos descomponer la distancia de un punto a la media como la distancia del punto a la recta + la distancia de la recta a la media.

Las distancias de todos los puntos a la recta azul (la media) es:

$$\sum_i (y_i - \bar{y})^2$$

La cual es llamada **Suma de Cuadrados Total** (SCT). Al sumar y restar $\hat{y}_i$ (predicción de la recta de regresión):

$$\begin{aligned} SCT &= \sum_i (y_i - \bar{y})^2 \\ &= \sum_i (y_i - \hat{y}_i + \hat{y}_i - \bar{y})^2 \\ &= \sum_i [(y_i - \hat{y}_i) + (\hat{y}_i - \bar{y})]^2 \\ &= \sum_i (y_i - \hat{y}_i)^2 + \sum_i (\hat{y}_i - \bar{y})^2 \end{aligned}$$

En el último paso, desapareció el doble producto por razones que veremos más adelante.

Retomando el final de la diapositiva anterior:

$$\underbrace{\sum_i (y_i - \bar{y})^2}_{SCT} = \underbrace{\sum_i (y_i - \hat{y}_i)^2}_{SCRes} + \underbrace{\sum_i (\hat{y}_i - \bar{y}_i)^2}_{SCExp}$$

Donde:

- *SCRes*: Suma de Cuadrados de los **Residuos**.
- *SCExp*: Suma de Cuadrados **Explicada**.

Retomando el final de la diapositiva anterior:

$$\underbrace{\sum_i (y_i - \bar{y})^2}_{SCT} = \underbrace{\sum_i (y_i - \hat{y}_i)^2}_{SCRes} + \underbrace{\sum_i (\hat{y}_i - \bar{y}_i)^2}_{SCExp}$$

Donde:

- *SCRes*: Suma de Cuadrados de los **Residuos**.
- *SCExp*: Suma de Cuadrados **Explicada**.

Como de antemano *SCT* está fija, al aumentar *SCRes*, disminuye *SCExp* y viceversa. En una situación ideal queremos que *SCExp* sea alta y que *SCRes* sea baja.

A partir de la partición de las sumas de cuadrados se define el coeficiente de determinación R^2.

$$R^2 = \frac{SCExp}{SCT} = 1 - \frac{SCres}{SCT}$$

Este estadístico nos permite determinar qué porcentaje de la variabilidad de Y puede explicarse a través de las variables contenidas en la matriz X.

A partir de la partición de las sumas de cuadrados se define el coeficiente de determinación R^2.

$$R^2 = \frac{SCExp}{SCT} = 1 - \frac{SCres}{SCT}$$

Este estadístico nos permite determinar qué porcentaje de la variabilidad de Y puede explicarse a través de las variables contenidas en la matriz X.

Una propiedad interesante de este indicador es que en el modelo de RLS, se puede probar que $R^2 = r^2_{xy}$.

Coeficiente de determinación

A partir de la partición de las sumas de cuadrados se define el coeficiente de determinación R^2.

$$R^2 = \frac{SCExp}{SCT} = 1 - \frac{SCres}{SCT}$$

Este estadístico nos permite determinar qué porcentaje de la variabilidad de Y puede explicarse a través de las variables contenidas en la matriz X.

Una propiedad interesante de este indicador es que en el modelo de RLS, se puede probar que $R^2 = r^2_{xy}$.

Volvamos al ejemplo que estábamos viendo en el R para obtener el valor de este estadístico.

Ejercicio 2 - Práctico 1

Considere el modelo de regresión lineal simple dado por la ecuación:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \qquad \forall_i = 1, \ldots, n$$

A partir de la ecuación $\hat{\beta}_{MCO} = \left(X^{'}X\right)^{-1} X^{'}Y$ demuestre que:

1. El primer elemento del vector $\hat{\beta}_{MCO}$ es $\bar{y} - \hat{\beta}_1 \bar{x}$.
2. El segundo elemento del vector $\hat{\beta}_{MCO}$ es $\frac{\sum_i (y_i - \bar{y})(x_i - \bar{x})}{\sum_i (x_i - \bar{x})^2}$
3. $\sum_i y_i = \sum_i \hat{y}_i$
4. Definiendo los residuos $\hat{\varepsilon}_i = y_i - \hat{y}_i$, compruebe que $\sum_i \hat{\varepsilon}_i = 0$.
5. De manera más general, comprueba que el vector de residuos es ortogonal a todas las columnas de la matriz X.

Ejercicio 5 - Práctico 1

Mejores tiempos conseguidos en pruebas de velocidad en atletismo en los Juegos Olímipicos de Atlanta 96.

Distancia	Tiempo (hombres)	Tiempo (mujeres)
100	9,84	10,84
200	19,32	22,12
400	43,19	48,25
800	102,58	117,73
1500	215,78	240,83
5000	787,96	899,88
10000	1627,34	1861,63
42195	7956,00	8765,00

1. Ajuste dos modelos de RLS (uno para hombres y otro para mujeres) y grafique.
2. Obtenga la predicción (en cada modelo) en una carrera de 3000 metros.
3. Calcule el valor del R^2 en ambos casos e interprete su valor.

Empezaremos con el modelo de regresión lineal múltiple.

- Veremos la interpretación geométrica del método de MCO.
- Veremos el teorema de Gauss-Markov.
- Generalizaremos algunos resultados del modelo de RLS.
- Definiremos el vector de *predichos* y de *residuos*.

Carmona, Francesc (2003). *Modelos Lineales (notas de curso).* Departament d'Estadística.

Farraway, Julian (2014). *Linear Models with R, second edition.* Chapman Hall/CRC.

Rencher, Alvin y Bruce Schaalje (2008). *Linear Models in Statistics, second edition.* John Wiley Sons, Inc.

¿Preguntas?

Muchas Gracias