



Econometría I

El Modelo Lineal General (I): Estimación

Miguel Jerez y Sonia Sotoca

Universidad Complutense de Madrid

Septiembre 2009

Índice

- **El modelo lineal general**
- Hipótesis del modelo
- Mínimos cuadrados ordinarios
- Máxima verosimilitud
- Medidas de ajuste
- Anexos

El modelo lineal general (I): Definición

Sea el Modelo Lineal General (MLG), definido por:

$$y_t = \sum_{i=1}^k \beta_i x_{ti} + \varepsilon_t \quad (t = 1, 2, \dots, n) \quad (1)$$

en donde:

y_t : observación t -ésima de la variable endógena o dependiente,

x_{ti} : observación t -ésima de la i -ésima variable exógena, variable explicativa o regresor,

β_i : i -ésimo parámetro o coeficiente,

ε_t : t -ésimo valor del término de error o perturbación,

k : número de parámetros, y

n : número de observaciones o casos de la muestra.

Por tanto el MLG define una relación:

- lineal entre una variable endógena y k variables explicativas,
- estocástica, ya que admite errores de ajuste, y
- útil para inferir los valores y_t condicionados a x_{ti} ($i = 1, 2, \dots, k$)

El modelo lineal general (II): Formulaciones matriciales

En notación vectorial, la expresión (1) puede escribirse como:

$$y_t = \mathbf{x}_t^T \boldsymbol{\beta} + \varepsilon_t \quad (t = 1, 2, \dots, n)$$

en donde:

\mathbf{x}_t^T : vector (1xk) de observaciones de cada una de las k variables explicativas correspondientes al caso t -ésimo, y

$\boldsymbol{\beta}$: vector (kx1) de parámetros.

o, de forma más compacta, como:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (2)$$

en donde:

\mathbf{y} : vector (nx1) de observaciones de la variable endógena, y

\mathbf{X} : matriz (nxk) que recoge en cada fila las observaciones de todas las variables explicativas correspondientes a cada valor de la variable endógena y, en cada columna, recoge todas las observaciones de cada variable explicativa:

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1^T \\ \mathbf{x}_2^T \\ \vdots \\ \mathbf{x}_n^T \end{bmatrix} = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1k} \\ X_{21} & X_{22} & \cdots & X_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{nk} \end{bmatrix}$$

El modelo lineal general (III): Interpretación de los coeficientes

Cuando las variables explicativas son continuas, los coeficientes de un modelo de regresión pueden interpretarse como derivadas (parciales) de la variable endógena con respecto a las variables explicativas. Si las variables tienen algún tipo de transformación, esta interpretación general puede concretarse de varias formas. Por ejemplo:

Modelo	Interpretación matemática y conceptual	
$y_t = \beta x_t + \varepsilon_t$	$\beta = \frac{dy_t}{dx_t}$	Cambio esperado en y_t cuando x_t aumenta en una unidad
$\ln y_t = \beta \ln x_t + \varepsilon_t$	$\beta = \frac{d \ln y_t}{d \ln x_t} = \frac{x_t}{y_t} \frac{dy_t}{dx_t}$	Cambio porcentual (en tanto por uno) esperado en y_t cuando x_t aumenta un uno por ciento (en tanto por uno)
$\ln y_t = \beta x_t + \varepsilon_t$	$\beta = \frac{d \ln y_t}{dx_t} = \frac{1}{y_t} \frac{dy_t}{dx_t}$	Cambio porcentual (en tanto por uno) esperado en y_t cuando x_t aumenta en una unidad
$y_t = \beta \ln x_t + \varepsilon_t$	$\beta = \frac{dy_t}{d \ln x_t} = x_t \frac{dy_t}{dx_t}$	Cambio esperado en y_t cuando x_t aumenta un uno por ciento (en tanto por uno)

El modelo lineal general (IV): Utilidad

Entre otros usos, los modelos de regresión resultan útiles para:

- **Predecir** los valores de las variables endógenas que con mayor probabilidad acompañarán a un conjunto de valores concretos de las variables explicativas. Ejemplo: Los sistemas de *scoring* crediticio de los bancos comerciales estiman la probabilidad de impago de las solicitudes de préstamo en función de las características de la operación y del solicitante
- **Controlar**, calculando los valores las variables explicativas sobre las que se tiene influencia que generarán con mayor probabilidad el valor que se desea para la variable endógena. Ejemplo: Los bancos centrales calculan el nivel de los tipos de intervención que llevará la tasa esperada de inflación al objetivo
- **Descomponer**, estimando la contribución de cada variable explicativa a los valores observados de las variables endógenas. Ejemplo: Calcular la contribución de la inversión publicitaria a las ventas, dados el precio de un producto
- **Simular** el comportamiento de la variable endógena ante: (a) una hipótesis sobre el comportamiento de las variables exógenas y (b) valores aleatorios del término de error. Por ejemplo, se desea simular el rendimiento de una cartera de valores ante una hipótesis de revalorización de un índice bursátil para calcular numéricamente la probabilidad de pérdidas
- **Estimar** una derivada o elasticidad. Por ejemplo, el cambio de valor esperado en una cartera ante movimientos en un índice es un parámetro importante para decidir la cobertura óptima de la cartera

El modelo lineal general (V): Cuestiones abiertas

El MLG plantea, inicialmente, tres problemas estadísticos relevantes:

- **Estimación**, que consiste en obtener una buena aproximación al valor de los parámetros a partir de una muestra de las variables x_t e y_t .
- **Inferencia** (o **contraste de hipótesis**), acerca del verdadero valor de los parámetros.
- **Previsión** de valores no observados de la variable endógena a partir del modelo estimado y los correspondientes valores de las variables exógenas.

Para resolver estos tres problemas es necesario hacer una serie de hipótesis acerca del MLG. Esto da lugar un cuarto problema:

- **Diagnosis** que consiste en, una vez estimado el modelo tentativo:
 - detectar incumplimientos de las hipótesis,
 - valorar sus posibles efectos negativos sobre la estimación, inferencia y previsión y, si se considera necesario,
 - resolver los problemas que puedan producirse.

Índice

- El modelo lineal general
- **Hipótesis del modelo**
- Mínimos cuadrados ordinarios
- Máxima verosimilitud
- Medidas de ajuste
- Anexos

Hipótesis del modelo (I)

[H.1] El modelo está correctamente especificado. Dentro de esta hipótesis general, pueden distinguirse tres hipótesis parciales:

[H.1.1] La relación entre la variable endógena y las variables explicativas es lineal.

[H.1.2] El modelo incluye todas las variables explicativas relevantes.

[H.1.3] El modelo no incluye ninguna variable explicativa irrelevante.

[H.2] Los parámetros del modelo son constantes, al menos dentro de la muestra que se está considerando.

[H.3] Suficientes grados de libertad. El número de observaciones es al menos igual que el número de parámetros que se desea estimar. Matemáticamente: $n \geq k$.

[H.4] Regresores no estocásticos. Las variables explicativas son deterministas.

[H.5] Ausencia de colinealidad. Las variables explicativas son linealmente independientes o, equivalentemente, que $|\mathbf{X}^T \mathbf{X}| \neq 0$

[H.6] Perturbaciones esféricas:

[H.6.1] Esperanza nula: $E(\varepsilon_t) = 0$ ($t = 1, 2, \dots, n$)

[H.6.2] Homoscedasticidad: $\text{var}(\varepsilon_t) = E(\varepsilon_t^2) = \sigma^2$ ($t = 1, 2, \dots, n$)

[H.6.3] Ausencia de autocorrelación: $\text{COV}(\varepsilon_t, \varepsilon_\tau) = E(\varepsilon_t \varepsilon_\tau) = 0$ ($t \neq \tau$)

Hipótesis del modelo (II)

[H.7] Normalidad. La distribución de probabilidad del término de error es normal.

A menudo las hipótesis [H.6] y [H.7] se resumen en el siguiente enunciado:

“El término de error del MLG se distribuye idéntica e independientemente como una variable aleatoria normal, de media nula y varianza constante”.

En notación matemática:

$$\varepsilon | \mathbf{X} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}) \quad \text{o bien:} \quad \varepsilon_t | \mathbf{x}_t^T \sim iid N(0, \sigma^2) \quad (t = 1, 2, \dots, n)$$

Teniendo en cuenta la expresión (2) resulta trivial demostrar que:

$$\mathbf{y} | \mathbf{X} \sim N(\mathbf{X} \boldsymbol{\beta}, \sigma^2 \mathbf{I}) \quad \text{o bien:} \quad y_t | \mathbf{x}_t^T \sim iid N(\mathbf{x}_t^T \boldsymbol{\beta}, \sigma^2) \quad (t = 1, 2, \dots, n)$$

Por tanto, las hipótesis [H.1], [H.4], [H.6] y [H.7] permiten caracterizar completamente la estructura estocástica de \mathbf{y} dada la información disponible en \mathbf{X} .

Índice

- El modelo lineal general
- **Hipótesis del modelo**
- **Mínimos cuadrados ordinarios**
- Máxima verosimilitud
- Medidas de ajuste
- Anexos

Mínimos cuadrados ordinarios (I): Estimador

Lo que caracteriza a un método de estimación es:

- el criterio que utiliza para estimar los parámetros β a partir de la muestra, y
- sus propiedades frente a otros estimadores.

A partir del MLG en forma vectorial [expresión (2)], y una estimación concreta de β , que denotaremos $\hat{\beta}$, definimos:

- el vector de valores ajustados o “previsiones intramuestrales”, $\hat{\mathbf{y}} = \mathbf{X} \hat{\beta}$
- el correspondiente vector de residuos como: $\hat{\boldsymbol{\varepsilon}} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{X} \hat{\beta}$

Un posible criterio consiste en calcular el valor de $\hat{\beta}$ que minimiza la suma de los residuos al cuadrado. Este criterio se conoce como de **mínimos cuadrados ordinarios** (MCO).

Para aplicarlo es necesario resolver el siguiente problema de optimización:

$$\text{Min}_{\hat{\beta}} S(\hat{\beta}) = (\mathbf{y} - \mathbf{X} \hat{\beta})^T (\mathbf{y} - \mathbf{X} \hat{\beta}) = \mathbf{y}^T \mathbf{y} - 2 \hat{\beta}^T \mathbf{X}^T \mathbf{y} + \hat{\beta}^T \mathbf{X}^T \mathbf{X} \hat{\beta}$$

Cuyas condiciones de primer orden proporcionan la siguiente solución (bajo [H.3] y [H.5]):

$$\begin{aligned} \frac{\delta S(\hat{\beta})}{\delta \hat{\beta}^T} = 0; \quad -2 \mathbf{X}^T \mathbf{y} + 2 \mathbf{X}^T \mathbf{X} \hat{\beta}_{\text{MCO}} = 0; \\ \mathbf{X}^T \mathbf{X} \hat{\beta}_{\text{MCO}} = \mathbf{X}^T \mathbf{y} \Leftrightarrow \hat{\beta}_{\text{MCO}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \end{aligned} \quad (3)$$

Mínimos cuadrados ordinarios (II): Normalidad e insesgadez

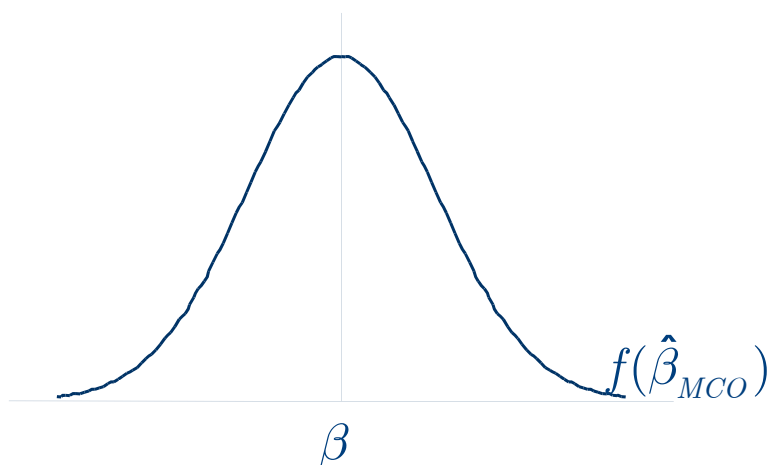
De acuerdo con [H.1], la expresión (3) puede escribirse como:

$$\hat{\beta}_{MCO} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T (\mathbf{X} \beta + \varepsilon) = \beta + (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \varepsilon \quad (4)$$

y, consecuentemente, si se cumplen [H.4], [H.6] y [H.7]:

- **la distribución del estimador es normal**, ya que $\hat{\beta}_{MCO}$ es una función lineal determinista de una variable aleatoria normal, y
- **el estimador es insesgado**; Aplicando el operador esperanza a ambos lados de (4) se obtiene $E(\hat{\beta}_{MCO} | \mathbf{X}) = \beta + (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T E(\varepsilon | \mathbf{X}) = \beta$

Por tanto, β es un vector determinista, pero su estimador por MCO es un vector de variables aleatorias normales, centradas en el valor que se quiere estimar.



Cada estimación de β es una muestra de la variable $\hat{\beta}_{MCO}$. La insesgadez significa que esta muestra probablemente saldrá del entorno del centro de la distribución, que coincide con el verdadero valor.

Mínimos cuadrados ordinarios (III): Eficiencia

Para caracterizar completamente la distribución del estimador, es necesario obtener su matriz de covarianzas. A partir de (4):

$$\begin{aligned} \text{cov}(\hat{\beta}_{MCO}) &= E\left[(\hat{\beta}_{MCO} - \beta)(\hat{\beta}_{MCO} - \beta)^T \mid \mathbf{X}\right] = \\ &= E\left[(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \varepsilon \varepsilon^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1}\right] = \\ &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T E(\varepsilon \varepsilon^T \mid \mathbf{X}) \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} = \sigma_\varepsilon^2 (\mathbf{X}^T \mathbf{X})^{-1} \end{aligned} \quad (5)$$

Teorema (Gauss-Markov): Si se cumplen las hipótesis [H.1]-[H.6] del MLG, entonces

$$\text{cov}(\bar{\beta}) - \text{cov}(\hat{\beta}_{MCO})$$

es una matriz semidefinida positiva, siendo $\bar{\beta}$ cualquier estimador lineal e insesgado de β

- De forma sintética, este teorema puede enunciarse diciendo que el estimador MCO es “BLUE” (**B**est **L**inear **U**nbiased **E**stimator)
- Este resultado es válido aunque los errores no sigan una distribución normal.
- Si los errores son normales, la estimación MCO es máximo verosímil y óptima dentro de la familia de estimadores insesgados de β

Este teorema se demuestra en el Anexo A.1.

Mínimos cuadrados ordinarios (IV): Estimación de la varianza residual y la matriz de covarianzas

Como vimos en (5), $\text{cov}(\hat{\beta}_{MCO}) = \sigma_{\varepsilon}^2 (\mathbf{X}^T \mathbf{X})^{-1}$. Evidentemente esta expresión no es computable ya que el verdadero valor de σ_{ε}^2 es, en general desconocido.

Para estimar la varianza del término de error puede usarse la expresión:

$$\hat{\sigma}_{\varepsilon}^2 = \frac{1}{n-k} \hat{\varepsilon}^T \hat{\varepsilon} \quad (6)$$

Que proporciona estimaciones insesgadas (ver Anexo A.2).

Consecuentemente, la matriz de covarianzas del estimador MCO puede estimarse usando la expresión:

$$\hat{\text{cov}}(\hat{\beta}_{MCO}) = \hat{\sigma}_{\varepsilon}^2 (\mathbf{X}^T \mathbf{X})^{-1} \quad (7)$$

y, si $E(\hat{\sigma}_{\varepsilon}^2) = \sigma_{\varepsilon}^2$ y las variables explicativas son números fijos, esta estimación será insesgada.

Mínimos cuadrados ordinarios (V): Distribución de los valores ajustados y los residuos

A partir de (4) y de la definición de los valores ajustados, resulta inmediato que:

$$\hat{\mathbf{y}} = \mathbf{X} \left[\boldsymbol{\beta} + (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\varepsilon} \right] = \mathbf{X} \boldsymbol{\beta} + \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\varepsilon}$$

y, consecuentemente, bajo las hipótesis [H.1], [H.4], [H.6] y [H.7] del MLG, resulta:

$$E(\hat{\mathbf{y}} | \mathbf{X}) = \mathbf{X} \boldsymbol{\beta}$$

$$\text{cov}(\hat{\mathbf{y}} | \mathbf{X}) = \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T E(\boldsymbol{\varepsilon} \boldsymbol{\varepsilon}^T) \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T = \sigma_\varepsilon^2 \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$$

luego: $\hat{\mathbf{y}} | \mathbf{X} \sim N(\mathbf{X} \boldsymbol{\beta}, \sigma_\varepsilon^2 \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T)$

Por otra parte,

$$\hat{\boldsymbol{\varepsilon}} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\varepsilon} - \mathbf{X} \boldsymbol{\beta} - \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\varepsilon} = \left[\mathbf{I} - \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \right] \boldsymbol{\varepsilon}$$

Por tanto, bajo las hipótesis del MLG podemos escribir:

$$E(\hat{\boldsymbol{\varepsilon}} | \mathbf{X}) = \mathbf{0}$$

$$\begin{aligned} \text{cov}(\hat{\boldsymbol{\varepsilon}} | \mathbf{X}) &= \left[\mathbf{I} - \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \right] E(\boldsymbol{\varepsilon} \boldsymbol{\varepsilon}^T) \left[\mathbf{I} - \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \right] \\ &= \sigma_\varepsilon^2 \left[\mathbf{I} - \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \right] \left[\mathbf{I} - \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \right] = \sigma_\varepsilon^2 \left[\mathbf{I} - \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \right] \end{aligned}$$

luego: $\hat{\boldsymbol{\varepsilon}} | \mathbf{X} \sim N\left\{ \mathbf{0}, \sigma_\varepsilon^2 \left[\mathbf{I} - \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \right] \right\}$

Mínimos cuadrados ordinarios (VI): Propiedades algebraicas

Cuando se estima el MLG por MCO, se cumplen las siguientes propiedades:

$$\mathbf{X}^T \hat{\boldsymbol{\varepsilon}} = \mathbf{0} \quad (8)$$

$$\hat{\mathbf{y}}^T \hat{\boldsymbol{\varepsilon}} = 0 \quad (9)$$

$$\mathbf{y}^T \mathbf{y} = \hat{\mathbf{y}}^T \hat{\mathbf{y}} + \hat{\boldsymbol{\varepsilon}}^T \hat{\boldsymbol{\varepsilon}} \quad (10)$$

y, si el modelo tiene término constante se cumplen, además, las siguientes propiedades:

$$\mathbf{i}_n^T \mathbf{y} = \mathbf{i}_n^T \hat{\mathbf{y}} \quad (11)$$

$$\mathbf{i}_n^T \hat{\boldsymbol{\varepsilon}} = 0 \quad (12)$$

$$(\mathbf{y} - \hat{\boldsymbol{\mu}}_y)^T (\mathbf{y} - \hat{\boldsymbol{\mu}}_y) = (\hat{\mathbf{y}} - \hat{\boldsymbol{\mu}}_{\hat{\mathbf{y}}})^T (\hat{\mathbf{y}} - \hat{\boldsymbol{\mu}}_{\hat{\mathbf{y}}}) + \hat{\boldsymbol{\varepsilon}}^T \hat{\boldsymbol{\varepsilon}} \quad (13)$$

siendo \mathbf{i}_n un vector ($n \times 1$) cuyas componentes son todas iguales a la unidad y $\hat{\boldsymbol{\mu}}_x$ un vector ($n \times 1$) cuyas componentes son todas iguales a la media muestral de la variable que aparece en el subíndice.

Por último, las propiedades (11)-(13) pueden expresarse de forma alternativa como:

$$\hat{\boldsymbol{\mu}}_y = \hat{\boldsymbol{\mu}}_{\hat{\mathbf{y}}} \quad (11')$$

$$\hat{\boldsymbol{\mu}}_{\hat{\boldsymbol{\varepsilon}}} = 0 \quad (12')$$

$$\hat{\sigma}_y^2 = \hat{\sigma}_{\hat{\mathbf{y}}}^2 + \hat{\sigma}_{\hat{\boldsymbol{\varepsilon}}}^2 \quad (13')$$

Índice

- El modelo lineal general
- Hipótesis del modelo
- Mínimos cuadrados ordinarios
- **Máxima verosimilitud**
- Medidas de ajuste
- Anexos

Máxima verosimilitud

Como vimos en apartados anteriores, bajo las hipótesis habituales se cumple que:

$$\mathbf{y} | \mathbf{X} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I})$$

Por tanto la función de verosimilitud de las estimaciones, condicionada a la muestra es:

$$L(\hat{\boldsymbol{\beta}}, \hat{\sigma}^2 | \mathbf{y}, \mathbf{X}) = (2\pi)^{-n/2} (\hat{\sigma}^2)^{-n/2} \exp \left\{ -\frac{1}{2\hat{\sigma}^2} (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^T (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \right\}$$

y, consecuentemente, su logaritmo es:

$$\ell(\hat{\boldsymbol{\beta}}, \hat{\sigma}^2 | \mathbf{y}, \mathbf{X}) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\hat{\sigma}^2) - \frac{1}{2\hat{\sigma}^2} (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^T (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$$

y las condiciones necesarias de primer orden para maximizar esta función son:

$$\frac{\partial \ell(\cdot)}{\partial \hat{\boldsymbol{\beta}}^T} = 0; -\frac{1}{2\hat{\sigma}^2} \frac{\partial}{\partial \hat{\boldsymbol{\beta}}^T} \left\{ (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^T (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \right\} = 0; \hat{\boldsymbol{\beta}}_{MV} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$$

$$\frac{\partial \ell(\cdot)}{\partial \hat{\sigma}^2} = 0; -\frac{n}{2} \frac{1}{\hat{\sigma}^2} + \frac{1}{2\hat{\sigma}^4} (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^T (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = 0; \hat{\sigma}_{MV}^2 = \frac{1}{n} \hat{\boldsymbol{\varepsilon}}^T \hat{\boldsymbol{\varepsilon}}$$

Teniendo en cuenta resultados anteriores, el estimador MV: (a) de $\boldsymbol{\beta}$ coincide con el MCO y (b) de la varianza del error es sesgado.

Índice

- El modelo lineal general
- Hipótesis del modelo
- Mínimos cuadrados ordinarios
- Máxima verosimilitud
- **Medidas de ajuste**
- Anexos

Medidas de ajuste (I): El coeficiente de determinación

Las medidas de ajuste sirven para:

- cuantificar la reducción de incertidumbre que proporciona el modelo y
- comparar modelos alternativos para la misma muestra.

La medida de ajuste más conocida es el coeficiente de determinación o R^2 . Este estadístico mide el porcentaje de la varianza de la variable dependiente que explica el modelo. El coeficiente de determinación se define como:

$$R^2 = \frac{(\hat{\mathbf{y}} - \hat{\mu}_{\hat{\mathbf{y}}})^T (\hat{\mathbf{y}} - \hat{\mu}_{\hat{\mathbf{y}}})}{(\mathbf{y} - \hat{\mu}_{\mathbf{y}})^T (\mathbf{y} - \hat{\mu}_{\mathbf{y}})} = \frac{\text{vâr}(\hat{\mathbf{y}})}{\text{vâr}(\mathbf{y})} \quad (14)$$

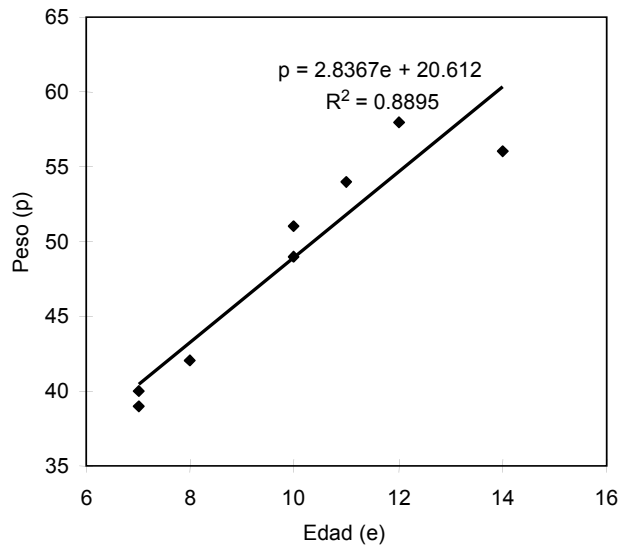
o, si el modelo tiene término constante:

$$R^2 = 1 - \frac{\hat{\boldsymbol{\varepsilon}}^T \hat{\boldsymbol{\varepsilon}}}{(\mathbf{y} - \hat{\mu}_{\mathbf{y}})^T (\mathbf{y} - \hat{\mu}_{\mathbf{y}})} = 1 - \frac{\text{vâr}(\hat{\boldsymbol{\varepsilon}})}{\text{vâr}(\mathbf{y})} \quad (15)$$

En principio, un modelo es tanto mejor cuanto mayor sea su correspondiente R^2 , ya que un valor alto supone que el modelo explica gran parte de la variabilidad de \mathbf{y} .

Medidas de ajuste (II): Ajuste y sobreajuste

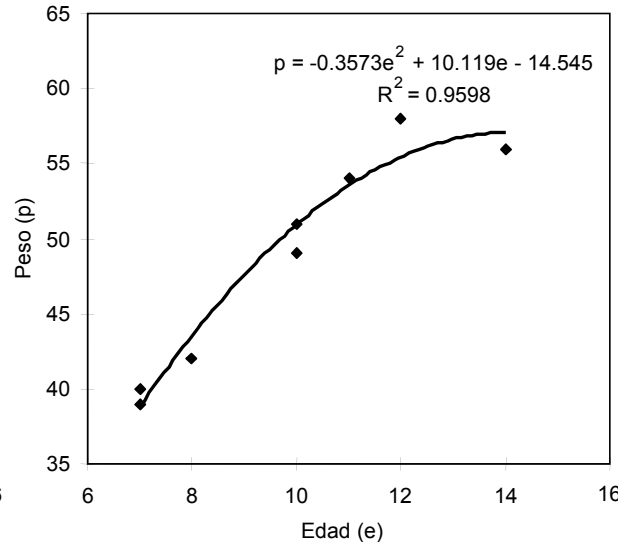
Las siguientes figuras muestran el ajuste de tres modelos distintos a una muestra que recoge el peso y la edad de un grupo de ocho niños. Como puede verse, el mayor R^2 no siempre corresponde al mejor modelo.



Una regresión lineal explica el 89% de la varianza del peso

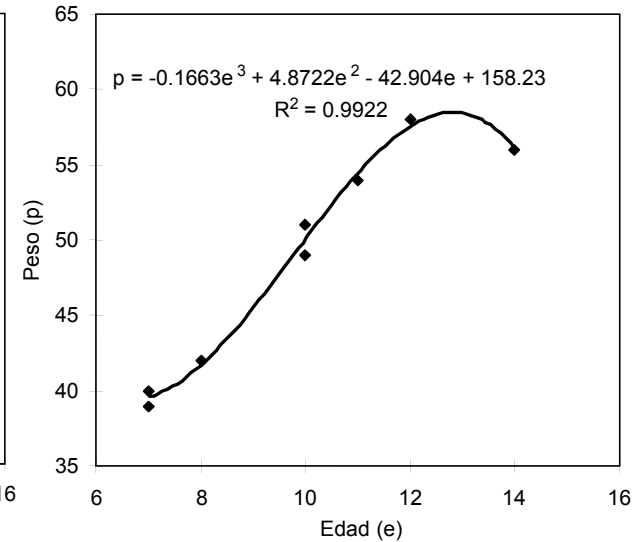
El modelo es imperfecto (¿qué peso predice para un niño de 0 años?)

Estas imperfecciones pueden deberse a: (a) limitaciones de la muestra y (b) no linealidad de la relación



Una regresión cuadrática mejora el R^2 (96%) y muestra un perfil intuitivamente razonable

A cambio, el nuevo modelo es más complejo, ya que requiere estimar tres parámetros en vez de dos



Una regresión cúbica proporciona un ajuste de más del 99%

Este es un buen resultado, siempre que estemos dispuestos a aceptar que los niños adelgazan a partir de los 13 años

Medidas de ajuste (III): Medidas alternativas

El uso mecánico del R^2 induce a sobreajustar la muestra. Para resolver este problema, a veces se usa un estadístico alternativo: el R^2 corregido con grados de libertad:

$$\bar{R}^2 = 1 - \frac{\frac{\hat{\varepsilon}^T \hat{\varepsilon}}{n-k}}{\frac{(\mathbf{y} - \hat{\mu}_y)^T (\mathbf{y} - \hat{\mu}_y)}{n-1}} = 1 - \frac{n-1}{n-k} (1 - R^2) \quad (16)$$

Este coeficiente: a) compara los estimadores insesgados de la varianza residual y de la varianza de la variable dependiente y b) penaliza los modelos con un elevado número de parámetros.

Actualmente disponemos de medidas más sofisticadas para comparar modelos, como por ejemplo los criterios de información de Akaike (AIC) y Schwartz (SBC)

$$AIC = n \ln(2\pi) + n \ln\left(\frac{\hat{\varepsilon}^T \hat{\varepsilon}}{n}\right) + n + 2(k+1) \quad (17)$$

$$SBC = n \ln(2\pi) + n \ln\left(\frac{\hat{\varepsilon}^T \hat{\varepsilon}}{n}\right) + n + (k+1) \ln(n) \quad (18)$$

El AIC prima la capacidad predictiva del modelo y tiende a sobreparametrizarlo, el SBC prima la especificación correcta.

Índice

- El modelo lineal general
- Hipótesis del modelo
- Mínimos cuadrados ordinarios
- Máxima verosimilitud
- Medidas de ajuste
- **Anexos**

Anexo A.1. Demostración del teorema de Gauss-Markov

Cualquier estimador lineal de β puede expresarse como:

$$\bar{\beta} = [(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T + \mathbf{D}] \mathbf{y} = [(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T + \mathbf{D}] (\mathbf{X} \beta + \varepsilon)$$

donde \mathbf{D} es una matriz ($k \times n$) arbitraria, que establece la diferencia entre el estimador MCO y el estimador alternativo. Desarrollando la igualdad anterior y tomando esperanzas:

$$E(\bar{\beta} | \mathbf{X}) = \beta + \mathbf{D} \mathbf{X} \beta; E(\bar{\beta} | \mathbf{X}) = \beta \Leftrightarrow \mathbf{D} \mathbf{X} = \mathbf{0}$$

por lo que el estimador alternativo sólo será insesgado si la matriz de distancia es ortogonal a las variables explicativas. A continuación obtenemos la matriz de covarianzas de este estimador.

$$\begin{aligned} \bar{\beta} &= \beta + \mathbf{D} \mathbf{X} \beta + [(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T + \mathbf{D}] \varepsilon \quad \dots \text{y, como } \mathbf{D} \mathbf{X} = \mathbf{0} \\ &= \beta + [(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T + \mathbf{D}] \varepsilon; \bar{\beta} - \beta = [(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T + \mathbf{D}] \varepsilon \end{aligned}$$

por tanto:

$$\begin{aligned} \text{cov}(\bar{\beta}) &= E \left\{ [(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T + \mathbf{D}] \varepsilon \varepsilon^T [(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T + \mathbf{D}]^T \middle| \mathbf{X} \right\} \\ &= \sigma_\varepsilon^2 \left[(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T + \mathbf{D} \right] \left[\mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} + \mathbf{D}^T \right] \\ &= \sigma_\varepsilon^2 \left[(\mathbf{X}^T \mathbf{X})^{-1} + \mathbf{D} \mathbf{D}^T \right] \end{aligned}$$

y $\mathbf{D} \mathbf{D}^T$ es una matriz semidefinida positiva, que es el resultado que queríamos demostrar.

Anexo A.2. Insesgadez del estimador de la varianza del error

$$\begin{aligned} E(\hat{\sigma}_\varepsilon^2) &= \frac{1}{n-k} E(\hat{\varepsilon}^T \hat{\varepsilon} | \mathbf{X}) && \text{y, como : } \hat{\varepsilon} = [I - \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T] \varepsilon = \mathbf{M} \varepsilon \\ &= \frac{1}{n-k} E(\varepsilon^T \mathbf{M} \mathbf{M} \varepsilon | \mathbf{X}) && \text{por ser } \mathbf{M} \text{ idempotente,} \\ &= \frac{1}{n-k} E(\varepsilon^T \mathbf{M} \varepsilon | \mathbf{X}) = \frac{1}{n-k} E[tr(\varepsilon^T \mathbf{M} \varepsilon) | \mathbf{X}] = \frac{1}{n-k} E[tr(\mathbf{M} \varepsilon \varepsilon^T) | \mathbf{X}] \\ &= \frac{1}{n-k} tr[\mathbf{M} E(\varepsilon \varepsilon^T | \mathbf{X})] = \sigma_\varepsilon^2 \frac{1}{n-k} tr(\mathbf{M}) \\ &= \sigma_\varepsilon^2 && \text{ya que } tr(\mathbf{M}) = n - k \end{aligned}$$

Miguel Jerez (mjerez@ccee.ucm.es)

Sonia Sotoca (sotoca@ccee.ucm.es)

Departamento de Fundamentos del Análisis Económico II
(Economía Cuantitativa)

Facultad de Ciencias Económicas, UCM

Más materiales en:

<http://www.ucm.es/info/ecocuan/mjm/ctr1mj>

<http://econometriamj.blogspot.com/>