Ordinary Least Squares (OLS) and Intro to Python

Juan Felipe Acevedo Pérez

Econometría Avanzada I
Maestría en Ciencias Económicas
Universidad Nacional de Colombia

1S2023

Table of Contents

Tabla de contenido

Random Vectors and Matrices

A random vector or random matrix is a vector or matrix whose elements are random variables. Informally, a random variable is defined as a variable whose value depends on the outcome of a chance experiment. (Formally, a random variable is a function defined for each element of a sample space.) (Rencher & Schaalje, 2008).

Consider the multiple regression model:

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

We treat the x variables as constants, in which case we have two random vectors

$$\mathrm{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \quad \text{and} \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \boldsymbol{\varepsilon}_n \end{pmatrix}$$

The **mean** or expected value of y is defined as:

$$\mu = E(y) = \int_{-\infty}^{\infty} y\ f(y)dy$$

For a constant a,

$$E(ay) = aE(y)$$

The **variance** of a random variable y is defined as:

$$\sigma^2 = \text{var}(y) = E(y - \mu)^2,$$

If a is a constant,

$$\text{var}(ay) = a^2\, \text{var}(y) = a^2\sigma^2$$

Independence

We define the **covariance** as:

$$\sigma_{ij} = \operatorname{cov}\left(y_i, y_j\right) = E\left[\left(y_i - \mu_i\right)\left(y_j - \mu_j\right)\right]$$

Two random variables yi and yj are said to be independent if their joint density factors into the product of their marginal densities

$$f\left(y_i, y_j\right) = f_i\left(y_i\right) f_j\left(y_j\right)$$

From the definition of independence, we obtain the following properties

1. $E\left(y_i, y_j\right) = E\left(y_i\right) E\left(y_j\right)$ if y_i and y_j are independent.
2. $\sigma_{ij} = \operatorname{cov}\left(y_i, y_j\right) = 0$ if y_i and y_j are independent.

Mean vectors and Covariance Matrices for Random Vectors

The expected value of a $p \times 1$ random vector $\mathbf{y}$ is defined as the vector of expected values of the p random variables $y_1, y_2, \ldots, y_p$ in $\mathbf{y}$:

$$E(\mathbf{y}) = E\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{pmatrix} = \begin{pmatrix} E(y_1) \\ E(y_2) \\ \vdots \\ E(y_p) \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{pmatrix} = \boldsymbol{\mu},$$

The variances $\sigma_1^2, \sigma_2^2, \ldots, \sigma_p^2$ of $y_1, y_2, \ldots, y_p$ and the covariances σ_{ij} for all $i \neq j$ can be conveniently displayed in the covariance matrix, which is denoted by $\boldsymbol{\Sigma}$, the uppercase version of σ_{ij} :

$$\boldsymbol{\Sigma} = \text{cov}(\mathbf{y}) = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{pmatrix}$$

Theorems

Theorem 1.a Suppose that $\mathbf{y}$ is a random vector, $\mathbf{X}$ is a random matrix, $\mathbf{a}$ and $\mathbf{b}$ are vectors of constants, and $\mathbf{A}$ and $\mathbf{B}$ are matrices of constants. Then, assuming the matrices and vectors in each product are conformal, we have the following expected values:

(i) $E(\mathbf{Ay}) = \mathbf{A}E(\mathbf{y})$.

(ii) $E\left(\mathbf{a}'\mathbf{Xb}\right) = \mathbf{a}'E(\mathbf{X})\mathbf{b}$.

(iii) $E(\mathbf{AXB}) = \mathbf{A}E(\mathbf{X})\mathbf{B}$.

Theorem 1.b If $\mathbf{a}$ is a $p \times 1$ vector of constants and $\mathbf{y}$ is a $p \times 1$ random vector with covariance matrix $\boldsymbol{\Sigma}$, then the variance of $z = \mathbf{a}'\mathbf{y}$ is given by

$$\sigma_z^2 = \text{var}\left(\mathbf{a}'\mathbf{y}\right) = \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a}$$

Proof.

$$\begin{aligned}
\text{var}\left(\mathbf{a}'\mathbf{y}\right) &= E\left(\mathbf{a}'\mathbf{y} - \mathbf{a}'\boldsymbol{\mu}\right)^2 = E\left[\mathbf{a}'(\mathbf{y} - \boldsymbol{\mu})\right]^2 \\
&= E\left[\mathbf{a}'(\mathbf{y} - \boldsymbol{\mu})\mathbf{a}'(\mathbf{y} - \boldsymbol{\mu})\right] \\
&= E\left[\mathbf{a}'(\mathbf{y} - \boldsymbol{\mu})(\mathbf{y} - \boldsymbol{\mu})'\mathbf{a}\right] \\
&= \mathbf{a}'E\left[(\mathbf{y} - \boldsymbol{\mu})(\mathbf{y} - \boldsymbol{\mu})'\right]\mathbf{a} \\
&= \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a}
\end{aligned}$$

Theorem 1.c Let $u = \mathbf{x}'\mathbf{A}\mathbf{x}$, where $\mathbf{A}$ is a symmetric matrix of constants. Then

$$\frac{\partial u}{\partial \mathbf{x}} = \frac{\partial \left(\mathbf{x}'\mathbf{A}\mathbf{x}\right)}{\partial \mathbf{x}} = 2\mathbf{A}\mathbf{x}$$

Tabla de contenido

The Model

The Model:

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

or

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

The preceding three assumptions on ε_i or y_i can be expressed in terms of the model in:

1. $E(\boldsymbol{\varepsilon}) = \mathbf{0}$ or $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$.
2. $\operatorname{cov}(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{I}$ or $\operatorname{cov}(\mathbf{y}) = \sigma^2\mathbf{I}$.

Note that the assumption $\operatorname{cov}(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{I}$ includes both the previous assumptions $\operatorname{var}(\varepsilon_i) = \sigma^2$ and $\operatorname{cov}(\varepsilon_i, \varepsilon_j) = 0$.

Estimation

We seek $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k$ that minimize

$$\begin{aligned}\sum_{i=1}^{n} \hat{\varepsilon}_i^2 &= \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \\ &= \sum_{i=1}^{n} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \cdots - \hat{\beta}_k x_{ik} \right)^2 .\end{aligned}$$

To find the values of $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k$ that minimize $\sum_{i=1}^{n} \hat{\varepsilon}_i^2$, we could differentiate $\sum_i \hat{\varepsilon}_i^2$ with respect to each $\hat{\beta}_j$ and set the results equal to zero to yield $k+1$ equations that can be solved simultaneously for the $\hat{\beta}_j$'s.

Theorem 2.a If $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where $\mathbf{X}$ is $n \times (k+1)$ of rank $k+1 < n$, then the value of $\hat{\boldsymbol{\beta}} = \left(\hat{\beta}_0, \hat{\hat{\beta}}_1, \ldots, \hat{\boldsymbol{\beta}}_k\right)'$ that minimizes is

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

Proof.

$$\hat{\varepsilon}'\hat{\varepsilon} = \sum_{i=1}^{n}\left(y_i - \mathbf{x}_i'\hat{\beta}\right)^2 = (\mathbf{y} - \mathbf{X}\hat{\beta})'(\mathbf{y} - \mathbf{X}\hat{\beta})$$

where $\mathbf{x}_i' = (1, x_{i1}, \ldots, x_{ik})$ is the i th row of $\mathbf{X}$. When the product $(\mathbf{y} - \mathbf{X}\hat{\beta})'(\mathbf{y} - \mathbf{X}\hat{\beta})$ is expanded, two of the resulting four terms can be combined to yield [*property* : $b'c = c'b$]

$$\hat{\varepsilon}'\hat{\varepsilon} = \mathbf{y}'\mathbf{y} - 2\mathbf{y}'\mathbf{X}\hat{\beta} + \hat{\beta}'\mathbf{X}'\mathbf{X}\hat{\beta}$$

We can find the value of $\hat{\beta}$ that minimizes $\hat{\varepsilon}'\hat{\varepsilon}$ by differentiating $\hat{\varepsilon}'\hat{\varepsilon}$ with respect to $\hat{\beta}$ [using Theorem 1.c] and setting the result equal to zero:

$$\frac{\partial \hat{\varepsilon}'\hat{\varepsilon}}{\partial \hat{\beta}} = \mathbf{0} - 2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\hat{\beta} = \mathbf{0}$$

This gives the normal equations

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

If $\mathbf{X}$ is full-rank, $\mathbf{X}'\mathbf{X}$ is nonsingular, the solution is given by

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

Properties of the Least-Squares Estimator

Theorem 2.b. If $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$, then $\hat{\boldsymbol{\beta}}$ is an unbiased estimator for $\boldsymbol{\beta}$.

PROOF

$$
\begin{aligned}
E(\hat{\boldsymbol{\beta}}) &= E\left[\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y}\right] \\
&= \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'E(\mathbf{y}) \quad [\text{ by } (3.38)] \\
&= \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\beta} \\
&= \boldsymbol{\beta}.
\end{aligned}
$$

Theorem 2.c. If $\text{cov}(\mathbf{y}) = \sigma^2 \mathbf{I}$, the covariance matrix for $\hat{\beta}$ is given by $\sigma^2 (\mathbf{X}'\mathbf{X})^{-1}$.

PROOF

$$\begin{aligned}
\text{cov}(\hat{\beta}) &= \text{cov}\left[(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y} \right] \\
&= (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \, \text{cov}(\mathbf{y}) \left[(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \right]' \\
&= (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \left(\sigma^2 \mathbf{I}\right) \mathbf{X} (\mathbf{X}'\mathbf{X})^{-1} \\
&= \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{X} (\mathbf{X}'\mathbf{X})^{-1} \\
&= \sigma^2 (\mathbf{X}'\mathbf{X})^{-1}
\end{aligned}$$

Tabla de contenido

Introducción

Una **variable** es un espacio de memoria donde se almacena un dato, un espacio donde se guarda la información necesaria para realizar las acciones que ejecutan los programas.

Se compone de: el nombre (identificador) y el tipo de dato (Python es capaz de identificar los tipos de datos primarios cuando se codifican directamente, pero al traerlos de una forma exógena, como con el input, se debe especificar su tipo).

Datos primitivos

Enteros: se codifica como int.

Reales: se codifica como float.

Booleanos: se codifica como bool.

Caracteres: representan los símbolos definidos por el ASCII (American Standard Code for Information Interchange). Se codifica como str

Tipos de secuencias

Cadenas de caracteres: Secuencia de cero o más caracteres; se delimita con ' o “”; es inmutable (no puede ser cambiada).

Listas: Almacena datos heterogéneos, separándolos por una coma (,) y es delimitada por los paréntesis cuadrados [ ]. Son mutables.

Tuplas: Secuencia de datos separados por una coma (,), y delimitada por los paréntesis redondos. Son inmutables.

Diccionarios: Colección de parejas clave:valor donde los valores pueden ser recuperados principalmente por su clave. Se escribe separando los ítems por comas (,) y entre . La llave es inmutable pero el valor sí es mutable.

Teorema de Gauss-Markov

Los estimadores de MCO presentan una serie de propiedades estadísticas atractivas dados unos supuestos; tal afirmación se sustenta en el Teorema de Gauss-Markov:

Siguiendo a Gujarati y Porter (2010): "Dados los supuestos del modelo clásico de regresión lineal, los estimadores de mínimos cuadrados, dentro de la clase de estimadores lineales insesgados, tienen varianza mínima, es decir, son MELI" (p.72).

MELI (BLUE en inglés) hace referencia a Mejores Estimadores Lineales Insesgados.

Supuestos de Estructura

1. Número de observaciones mayor a número de parámetros a estimar.
2. Debe haber variación en los valores de las variables.
3. No debe haber colinealidad exacta entre las variables explicativas (correlación y VIF).
4. No hay sesgo de especificación (Test RESET de Ramsey o prueba CUSUM).

Supuestos del término de Error

1. Variables explicativas independientes del término de error.
2. Valor medio del término de error igual a cero.
3. Homoscedasticidad en el término de error. (Informalmente mediante gráfico y formalmente con Test BP o Test White)
4. No autocorrelación entre las perturbaciones. (Test Durbin-Watson, Ljung-Box, Box-Pierce o Breusch-Godfrey)

¿Supuesto de Normalidad?

NO es un supuesto del MCRL. Es un supuesto del Modelo Clásico de Regresión Lineal **Normal**.

- No se necesita para garantizar estimadores MELI. Se incluye para **validez en inferencia estadística.**
- Resumen: término de error normal e independientemente distribuido.

¿Cómo verificarlo?

- Formal:Jarque-Bera, Anderson-Darling, Kolmogorov-Smirnov.
- Informal: Examen visual de la distribución estimada de los residuos