CAPíTULO 3

El modelo clásico de regresión

En el capítulo anterior hemos aplicado el **algebra matricial** y la **estadística descriptiva** al modelo lineal general $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$ para encontrar el **estimador de mínimos cuadrados** ordinarios $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$. La teoría de matrices ha jugado un papel relevante en el desarrollo del tema: nos ha permitido ordenar el conjunto de datos en la matriz de diseño $\mathbf{X}$ y en el vector de observaciones $\mathbf{y}$, resolver el sistema de ecuaciones normales $\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}$ y establecer las **propiedades numéricas** de este método de estimación, $\mathbf{X}'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = \mathbf{X}'\hat{\mathbf{u}} = \mathbf{0}_k$. También hay que apreciar el papel jugado por la estadística descriptiva: nos revela que el estimador de mínimos cuadrados usa la información de los datos resumida en los momentos muestrales de primer y segundo orden $\sum_{h=1}^{n} X_{ih}$, $\sum_{h=1}^{n} X_{ih}X_{jh}$ y $\sum_{h=1}^{n} X_{ih}Y_h$, y nos sugiere medir la bondad del ajuste mediente el cudadrado de la correlación simple entre Y_i e $\hat{Y}_i$.

En este capítulo vamos a hacer uso de la **teoría de probabilidad** para estudiar las **propiedades estadísticas** del estimador de mínimos cuadrados. Vamos a especificar un conjunto de **supuestos básicos** bajo los cuales el estimador de mínimos cuadrados ordinarios es el *mejor* estimador que puede utilizarse porque cumple unas **propiedades estadísticas** deseables.

3.1. Supuestos básicos

Sea $\mathbf{y} = (Y_1\ Y_2\ \dots\ Y_n)'$ un vector de n-variables aleatorias y sea $\mathbf{X}$ una matriz $n \times k$ de variables explicativas. Suponemos que la esperanza matemática de $\mathbf{y}$ condicionada a $\mathbf{X}$, $E(\mathbf{y}|\mathbf{X})$, es una función lineal de un vector de parámetros $\boldsymbol{\beta} = (\beta_1\ \beta_2\ \dots\ \beta_k)'$, esto es,

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

y que el vector de variables aleatorias $\mathbf{y}$ puede representarse como

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \tag{3.1}$$

en donde $\mathbf{u} = (u_1\ u_2\ \dots\ u_n)'$ es un vector de n perturbaciones estocásticas.

Es conveniente interpretar la ecuación (3.1) como un **experimento estadístico** que puede repetirse en idénticas condiciones. Cada vez que se repite el experimento se obtiene un **resultado aleatorio**. El resultado del experimento representado por la ecuación (3.1) es un vector de observaciones. De aquí, los datos $\{y_1, y_2, \dots, y_n\}$ que se emplean en la estimación de un modelo de regresión se interpretan como una **realización** particular de las infinitas posibles realizaciones de una variable aleatoria n-dimensional $\{Y_1, Y_2, \dots, Y_n\}$. También se dice que los datos los datos $\{y_1, y_2, \dots, y_n\}$ son una **muestra** de la **población** $\{Y_1, Y_2, \dots, Y_n\}$. Para resaltar esta distinción entre muestra y población cualquier modelo estadístico y, en particular, el modelo de regresión se denomina también **proceso generador de datos**.

Observación 13. *En Econometría, es habitual utilizar la misma notación para las variables aleatorias* $\{Y_1, Y_2, \ldots, Y_n\}$ *y para los valores observados* $\{Y_1, Y_2, \ldots, Y_n\}$. *La notación, por tanto, es ambigua, pero la ambiguedad se resolverá en el contexto en que se utiliza.*

El modelo lineal general (3.1) cumple los **supuestos básicos** si:

1. $\mathbf{X}$ es una matriz no estocástica de rango $k < n$, tal que
$$\lim_{n\to\infty} \frac{\mathbf{X}'\mathbf{X}}{n} = \mathbf{Q}$$
siendo $\mathbf{Q}$ una matriz finita no singular (definida positiva) de orden $k \times k$,
2. $\mathbf{u}$ tiene una distribución normal multivariante con vector de medias nulo y matriz de varianzas y covarianzas escalar, $\mathbf{u} \sim N(\mathbf{0}, \sigma_u^2 \mathbf{I}_n)$

El significado de los supuestos referidos a la matriz de variables explicativas $\mathbf{X}$ es el siguiente:

1. Regresores no estocásticos. La matriz $\mathbf{X}$ es no estocástica cuando permanece fija en las diferentes repeticiones del experimento.
2. Ausencia de multicolinealidad. El rango de $\mathbf{X}$, $\rho(\mathbf{X}) = k$, es el número de columnas (o filas) linealmente independientes. Este supuesto implica que $\rho(\mathbf{X}'\mathbf{X}) = k$ y que el sistema de ecuaciones normales tiene solución única. Si el supuesto se incumple, $\rho(\mathbf{X}) < k$, entonces las columnas de la matriz $\mathbf{X}$ son linealmente dependientes, $\rho(\mathbf{X}'\mathbf{X}) < k$ y el sistema de ecuaciones normales tiene soluciones múltiples. El término *multicolinealidad* hace referencia a la existencia de una o más relaciones lineales exactas o perfectas entre las variables explicativas.
3. El supuesto $k < n$ indica que el número de observaciones es mayor que el número de parámetros a estimar. Si $k > n$, entonces $\rho(\mathbf{X}) \leq n$, $\rho(\mathbf{X}'\mathbf{X}) \leq n$, y el sistema de ecuaciones normales tendrá soluciones múltiples.
4. Momentos muestrales finitos. El elemento genérico de $\mathbf{X}'\mathbf{X}$ dividido por n es
$$\sum_{h=1}^{n} \frac{X_{ih}X_{jh}}{n}$$
que converge a una constante finita cuando $n \to \infty$.

En cuanto a los supuestos referidos al vector de perturbaciones $\mathbf{u}$,

1. Las perturbaciones estocásticas $u_i\,(i = 1, \ldots, n)$ tienen media cero, $E(u_i) = 0$.
2. Homocedasticidad. Las perturbaciones estocásticas $u_i\,(i = 1, \ldots, n)$ tienen la misma varianza, $V(u_i) = E[u_i - E(u_i)]^2 = E(u_i^2) = \sigma_u^2$. La notación $V(u_i) = \sigma_u^2$ indica que la varianza no cambia con el índice i. El incumplimiento de este supuesto se denomina heterocedasticidad, $V(u_i) = \sigma_i^2$.
3. Ausencia de autocorrelación o de correlación serial. Las perturbaciones estocásticas son mútuamente ortogonales: u_i y u_j tienen covarianza nula, $Cov(u_i, u_j) = E\{[u_i - E(u_i)][u_j - E(u_j)]\} = E(u_i u_j) = 0\ \forall i \neq j$. El incumplimiento de este supuesto se denomina autocorrelación, la covarianza $E(u_i u_j) \neq 0$ para algún $i \neq j$ (Nota: la correlación simple entre u_i y u_j es $E(u_i, u_j)/\sqrt{E(u_i^2)E(u_j^2)}$).
4. Normalidad. Las perturbaciones estocásticas $u_i\,(i = 1, \ldots, n)$ tienen una distribución normal, $u_i \sim N(0, \sigma_u^2)$.

Otra forma de resumir estas cuatro hipótesis es la siguiente: *los errores se distribuyen idéntica e independientemente como una normal con media cero y varianza constante* σ_u^2, $u_i \sim iidN(0, \sigma_u^2)$.

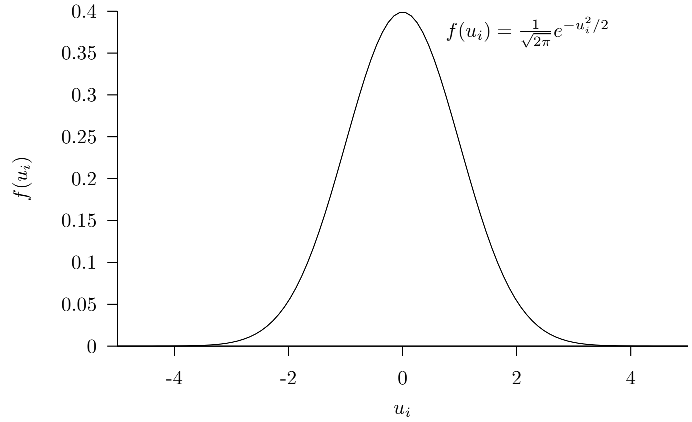

Figura 1: Función de densidad de probabilidad de la distribución normal estándar

El supuesto de que cada error u_i tiene media cero, $E(u_i)$, puede expresarse en forma matricial como

$$E(\mathbf{u}) = \begin{pmatrix} E(u_1) \\ E(u_2) \\ \vdots \\ E(u_n) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

Los supuestos de homocedasticidad y ausencia de autocorrelación implican que la matriz de varianzas y covarianzas del vector de perturbaciones $\mathbf{u}$ es escalar

$$V(\mathbf{u}) = E[(\mathbf{u} - E(\mathbf{u}))(\mathbf{u}' - E(\mathbf{u}'))] = E\left[\begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} \begin{pmatrix} u_1 & u_2 & \dots & u_n \end{pmatrix}\right]$$

$$= \begin{pmatrix} E(u_1^2) & E(u_1u_2) & \dots & E(u_1u_n) \\ E(u_2u_1) & E(u_2^2) & \dots & E(u_2u_n) \\ \vdots & \vdots & \ddots & \vdots \\ E(u_nu_1) & E(u_nu_2) & \dots & E(u_n^2) \end{pmatrix} = \begin{pmatrix} \sigma_u^2 & 0 & \dots & 0 \\ 0 & \sigma_u^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_u^2 \end{pmatrix} = \sigma_u^2 \mathbf{I}_n$$

Proposición 21. *Bajo los supuestos básicos, el vector de n-variables aleatorias* $\mathbf{y} = (Y_1 \; Y_2 \; \dots \; Y_n)'$ *en el modelo (3.1) tiene una distribución normal multivariante con vector de medias* $\mathbf{X}\boldsymbol{\beta}$ *y matriz de varianzas-covarianzas* $\sigma_u^2 I_n$,

$$\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma_u^2 I_n)$$

Demostración. En general, una combinación lineal de variables aleatorias independientes con distribución normal tiene también una distribución normal. Como $\mathbf{y}$ es una transformación lineal del vector $\mathbf{u}$, $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$, que tiene una distribución normal multivariante, $\mathbf{y}$ tiene también una distribución normal multivariante. El vector de medias de $\mathbf{y}$ es

$$E(\mathbf{y}) = E(\mathbf{X}\boldsymbol{\beta} + \mathbf{u}) = E(\mathbf{X}\boldsymbol{\beta}) + E(\mathbf{u}) = \mathbf{X}\boldsymbol{\beta}$$

y su matriz de varianzas y covarianzas

$$V(\mathbf{y}) = E\left[(\mathbf{y} - E(\mathbf{y}))(\mathbf{y} - E(\mathbf{y}))'\right] = E\left[(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\right] = E[\mathbf{u}\mathbf{u}'] = \sigma_u^2 I_n$$

□

Observación 14. *La distribución de probabilidad de la variable aleatoria* $\mathbf{y}$ *depende de los parámetros desconocidos* $\boldsymbol{\beta}$ *y* σ_u^2. *El método de estimación de mínimos cuadrados proporciona un estimador de* $\boldsymbol{\beta}$*; queda pendiente la estimación del parámetro* σ_u^2.

Definición 20. *La ecuación (3.1) se denomina función de* **regresión poblacional***; y la ecuación estimada, función de* **regresión muestral**.

Definición 21. *El modelo lineal general (3.1), junto con los supuestos sobre* $\mathbf{X}$ *y* $\mathbf{u}$*, excepto el de normalidad, se denomina* **modelo clásico de regresión**.

3.2. Estimador de σ_u^2

Las perturbaciones estocásticas $\{u_1, u_2, \ldots, u_n\}$ tienen varianza común σ_u^2. Si seleccionaramos una muestra $\{u_1, u_2, \ldots, u_n\}$, entonces podríamos estimar el parámetro poblacional σ_u^2 a partir de la varianza muestral

$$s_u^2 = \frac{\sum_{i=1}^n (u_i - \bar{u})^2}{n} = \frac{1}{n}\left[\mathbf{u}'\mathbf{u} - n\bar{u}^2\right]$$

donde $\bar{u} = \sum_{i=1}^n u_i/n$ es la media muestral. Ahora bien, como las perturbaciones u_i no son observables, el estimador s_u^2 no es calculable.

Para evitar este problema, podemos contemplar los residuos $\hat{u}_i$ como estimaciones de los errores u_i y estimar el parámetro σ_u^2 como la varianza muestral de los residuos. Suponiendo que el modelo de regresión tiene término constante,

$$\tilde{\sigma}_u^2 = \frac{\sum_{i=1}^n (\hat{u}_i - \bar{\hat{u}})^2}{n} = \frac{\sum_{i=1}^n \hat{u}_i^2}{n} = \frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{n}$$

que se denomina estimador de máxima verosimilitud de la varianza de las perturbaciones.

Alternativamente, y reconociendo que los grados de libertad de la suma de cuadrados de libertad son $n - k$, podemos proponer el estimador

$$\hat{\sigma}_u^2 = \frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{n-k} = \frac{\sum_{i=1}^n \hat{u}_i^2}{n-k}$$

que se denomina estimador de mínimos cuadrados de la varianza de las perturbaciones.

Definición 22. *La raíz cuadrada de* $\hat{\sigma}_u^2$, $\hat{\sigma}_u$, *se conoce como* **error estándar de la regresión**.

Ejemplo 1. En el modelo de las calificaciones, $n = 10$, $k = 4$ y la suma de cuadrados de los residuos $\mathbf{u}'\mathbf{u} = 6{,}7027$. De aquí, $\tilde{\sigma}_u^2 = 6{,}7027/10 = 0{,}67027$ y $\hat{\sigma}_u^2 = 6{,}7027/6 = 1{,}11712$.

◁

3.3. Propiedades estadísticas de $\hat{\boldsymbol{\beta}}$

El estimador $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ del vector de parámetros $\boldsymbol{\beta}$ es un **estadístico**, es decir, una función de la variable aleatoria n-dimensional $\{Y_1, Y_2, \ldots, Y_n\}$, $\hat{\boldsymbol{\beta}} : \Re^n \to \Re^k$. Para hacer explícita esta dependencia escribimos $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}(Y_1, Y_2, \ldots, Y_n)$. Una **estimación** es un valor específico del estimador calculado para una de las infinitas posibles realizaciones de la variable aleatoria $\{Y_1, Y_2, \ldots, Y_n\}$. Si $\{y_1, y_2, \ldots, y_n\}$ es una realización particular de la variable aleatoria $\{Y_1, Y_2, \ldots, Y_n\}$, entonces la estimación $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}(y_1, y_2, \ldots, y_n)$ es uno de los muchos posibles valores que puede tomar la variable aleatoria $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}(Y_1, Y_2, \ldots, Y_n)$.

La distribución de probabilidad conjunta del estimador $\hat{\boldsymbol{\beta}}(Y_1, Y_2, \ldots, Y_n)$ describe el comportamiento de las estimaciones que se obtendrían en el conjunto de posibles muestras de la población $\{Y_1, Y_2, \ldots, Y_n\}$. Esta distribución se denomina **distribución muestral** y puede derivarse de la distribución de probabilidad de $\{Y_1, Y_2, \ldots, Y_n\}$, $\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma_u^2\mathbf{I})$, que a su vez se ha derivado de la distribución de probabilidad de $\{u_1, u_2, \ldots, u_n\}$, $\mathbf{u} \sim N(\mathbf{0}, \sigma_u^2\mathbf{I})$.

Teorema 2. *Bajo los supuestos básicos, el estimador de mínimos cuadrados $\hat{\boldsymbol{\beta}}$ del vector de parámetros $\boldsymbol{\beta}$ en el modelo (3.1) tiene una distribución normal multivariante con vector de medias $\boldsymbol{\beta}$ y matriz de varianzas y covarianzas $\sigma_u^2(\mathbf{X}'\mathbf{X})^{-1}$, que se escribe sucintamente como*

$$\hat{\boldsymbol{\beta}} \sim N\left(\boldsymbol{\beta}, \sigma_u^2(\mathbf{X}'\mathbf{X})^{-1}\right)$$

Demostración.

1. Normalidad. Cada elemento $\hat{\beta}_j \, (j = 1, \ldots, k)$ del vector $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ es una combinación lineal de variables aleatorias independientes $Y_1, \ldots, Y_n$ con distribución normal,
$$\hat{\beta}_j = \sum_{i=1}^{n} c_i Y_i$$
en donde las ponderaciones $c_1, \ldots, c_n$ son los elementos de la fila j de la matriz $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$.
2. Vector de medias
$$E(\hat{\boldsymbol{\beta}}) = E\left[\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y}\right] = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'E\left[\mathbf{y}\right] = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\left[\mathbf{X}\boldsymbol{\beta}\right] = \boldsymbol{\beta}$$
3. Matriz de varianzas y covarianzas
$$V(\hat{\boldsymbol{\beta}}) = E\left[\left(\hat{\boldsymbol{\beta}} - E(\hat{\boldsymbol{\beta}})\right)\left(\hat{\boldsymbol{\beta}} - E(\hat{\boldsymbol{\beta}})\right)'\right]$$
Como $\hat{\boldsymbol{\beta}} - E(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\left[\mathbf{y} - E(\mathbf{y})\right]$, tenemos
$$\begin{aligned} V(\hat{\boldsymbol{\beta}}) =& E\left\{\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\left[\mathbf{y} - E(\mathbf{y})\right]\left[\mathbf{y} - E(\mathbf{y})\right]'\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\right\} \\ =& \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'E\left\{\left[\mathbf{y} - E(\mathbf{y})\right]\left[\mathbf{y} - E(\mathbf{y})\right]'\right\}\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1} \\ =& \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\left\{\sigma_u^2\mathbf{I}\right\}\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1} = \sigma_u^2\left(\mathbf{X}'\mathbf{X}\right)^{-1} \end{aligned}$$

□

Definición 23. *Un estimador $\hat{\beta}_i$ del parámetro β_i es **insesgado** si su esperanza matemática coincide con el verdadero parámetro β_i, $E(\hat{\beta}_i) = \beta_i$. En el caso multidimensional, un vector de estimadores $\hat{\boldsymbol{\beta}}$ es insesgado si $E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$.*

El Teorema 2 afirma que el estimador de mínimos cuadrados es insesgado: si tomamos diferentes muestras de tamaño n y para cada una calculamos el estimador $\hat{\boldsymbol{\beta}}$, entonces la media muestral de estas estimaciones es igual a $\boldsymbol{\beta}$.

Definición 24. *Un estimador insesgado $\hat{\beta}_i$ es más* ***eficiente*** *que otro estimador $\tilde{\beta}_i$ también insesgado, si la varianza muestral de $\hat{\beta}_i$ es menor que la de $\tilde{\beta}_i$, $V(\hat{\beta}_i) < V(\tilde{\beta}_i)$. En el caso multidimensional, un vector de estimadores insesgados $\hat{\boldsymbol{\beta}}$ es más eficiente que otro $\tilde{\beta}$, si la diferencia entre las matrices de varianzas y covarianzas $V(\hat{\boldsymbol{\beta}}) - V(\tilde{\boldsymbol{\beta}})$ es una matriz definida negativa.*

Observación 15. *Sea $\boldsymbol{\gamma} = \mathbf{w}'\boldsymbol{\beta}$ cualquier combinación lineal de los parámetros de $\boldsymbol{\beta}$. Entoces $\hat{\boldsymbol{\gamma}} = \mathbf{w}'\hat{\boldsymbol{\beta}}$ es más eficiente que $\tilde{\boldsymbol{\gamma}} = \mathbf{w}'\tilde{\boldsymbol{\beta}}$ si $V(\hat{\gamma}) < V(\tilde{\gamma})$, esto es, si*

$$\mathbf{w}'V(\hat{\boldsymbol{\beta}})\mathbf{w} - \mathbf{w}'V(\tilde{\boldsymbol{\beta}})\mathbf{w} = \mathbf{w}'\left[V(\hat{\boldsymbol{\beta}}) - V(\tilde{\boldsymbol{\beta}})\right]\mathbf{w}$$

es una forma cuadrática definida negativa.

La inversa de la varianza de un estimador es una medida de su precisión o acuracidad. Cuanto menor sea la varianza del estimador, tanto más preciso o acurado será el estimador, lo que significa que las estimaciones obtenidas en las distintas realizaciones del experimento aleatorio estarán próximas al parámetro que se desea estimar.

Teorema 3 (Teorema de Gauss-Markov). *Bajo los supuestos básicos del modelo clásico, el estimador de mínimos cuadrados $\hat{\boldsymbol{\beta}}$ es el más eficiente en la clase de estimadores lineales e insesgados de $\boldsymbol{\beta}$.*

Demostración. La clase general de estimadores lineales está definida por

$$\tilde{\beta} = \mathbf{C}\mathbf{y}$$

en donde $\mathbf{C}$ es una matriz de orden $k \times n$ de números fijos. Se observa que el estimador $\hat{\boldsymbol{\beta}}$ es un miembro particular de esta clase cuando $\mathbf{C} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$.

Dentro de la clase general de estimadores lineales, los estimadores insesgados

$$E(\tilde{\beta}) = E(\mathbf{C}\mathbf{y}) = \mathbf{C}\mathbf{X}\boldsymbol{\beta} = \boldsymbol{\beta}$$

son aquelos que cumplen $\mathbf{C}\mathbf{X} = \mathbf{I}_k$.

La matriz de varianzas y covarianzas de $\tilde{\beta}$ es

$$V(\tilde{\beta}) = E\left[\left(\tilde{\beta} - E(\tilde{\beta})\right)\left(\tilde{\beta} - E(\tilde{\beta})\right)'\right] = \mathbf{C}E\left[(\mathbf{y} - E(\mathbf{y}))(\mathbf{y} - E(\mathbf{y}))'\right]\mathbf{C}' = \sigma_u^2\mathbf{C}\mathbf{C}'$$

Ahora escribimos

$$\mathbf{C} = \mathbf{D} + \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'$$

en donde se cumple que $\mathbf{D}\mathbf{X} = 0$ porque $\mathbf{C}\mathbf{X} = \mathbf{I}_k$. De modo que

$$\mathbf{C}\mathbf{C}' = \left[\mathbf{D} + \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\right]\left[\mathbf{D}' + \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\right] = \mathbf{D}\mathbf{D}' + \left(\mathbf{X}'\mathbf{X}\right)^{-1}$$

Sustituyendo $\mathbf{C}\mathbf{C}'$ en $V(\tilde{\beta})$, tenemos

$$V(\tilde{\beta}) = \sigma_u^2\mathbf{D}\mathbf{D}' + \sigma_u^2\left(\mathbf{X}'\mathbf{X}\right)^{-1}$$

Esta ecuación puede escribirse como

$$V(\tilde{\beta}) - V(\hat{\boldsymbol{\beta}}) = \sigma_u^2\mathbf{D}\mathbf{D}'$$

donde vemos que la diferencia de las dos matrices de varianzas y covarianzas es una matriz semidefinida positiva. □

Observación 16. *El Teorema de Gauss-Markow no hace uso del supuesto de normalidad de las perturbaciones.*

Definición 25. *Un estimador $\hat{\beta}_i$ es consistente o converge en probabilidad al parámetro verdadero β_i si, para todo $\epsilon > 0$,*

$$\lim_{n\to\infty} P(|\hat{\beta}_i^{(n)} - \beta_i| \geq \epsilon) = 0$$

en donde $\hat{\beta}_i^{(n)}$ es el estimador calculado con n observaciones. En el caso multidimensional, el vector de estimadores $\hat{\boldsymbol{\beta}}$ del vector de parámetros $\boldsymbol{\beta}$ es consistente si, para todo $\epsilon > 0$,

$$\lim_{n\to\infty} P(\|\hat{\boldsymbol{\beta}}^{(n)} - \boldsymbol{\beta}\| \geq \epsilon) = 0$$

en donde $\hat{\boldsymbol{\beta}}^{(n)}$ es el vector de estimadores basado en una muestra de n observaciones y $\|\hat{\boldsymbol{\beta}}^{(n)} - \boldsymbol{\beta}\|$ es la norma euclídea del correspondiente vector.

En la definición anterior, β_i es el límite en probabilidad de la secuencia de variables aleatorias $\{\hat{\beta}_i^{(n)}\}_{n=k}^{\infty}$ y se escribe como

$$plim\hat{\beta}_i = \beta_i \quad \text{o} \quad \hat{\beta}_i \xrightarrow{p} \beta_i$$

Definición 26. *Un estimador $\hat{\beta}_i$ converge en media cuadrática al parámetro verdadero β_i si*

$$\lim_{n\to\infty} E(\hat{\beta}_i^{(n)} - \beta_i)^2 = 0$$

o, equivalentemente, si

$$\lim_{n\to\infty} sesgo(\hat{\beta}_i) \equiv \lim_{n\to\infty} \left[E(\hat{\beta}_i^{(n)}) - \beta_i\right] = 0 \quad y \quad \lim_{n\to\infty} var(\hat{\beta}_i^{(n)}) \equiv \lim_{n\to\infty} E(\hat{\beta}_i^{(n)} - \beta_i)^2 = 0$$

En el caso multidimensional, un vector de estimadores $\hat{\boldsymbol{\beta}}$ converge en media cuadrática al vector de parámetros verdaderos $\boldsymbol{\beta}$ si

$$\lim_{n\to\infty} E\left[(\hat{\boldsymbol{\beta}}^{(n)} - \boldsymbol{\beta})'(\hat{\boldsymbol{\beta}}^{(n)} - \boldsymbol{\beta})\right] = \lim_{n\to\infty} \sum_{i=1}^{k} E(\hat{\beta}_i^{(n)} - \beta_i)^2 = 0$$

Proposición 22. *Convergencia en media cuadrática implica convergencia en probabilidad.*

Proposición 23. *Bajo los supuestos básicos del modelo lineal general clásico, el estimador de mínimos cuadrados $\hat{\boldsymbol{\beta}}$ del vector de paramámetros $\boldsymbol{\beta}$ en el modelo (3.1) es consistente.*

Demostración. $\hat{\boldsymbol{\beta}}$ converge en media cuadrática a $\boldsymbol{\beta}$ (y, por la proposición 22, es consistente) porque es insesgado y su matriz de varianzas y covarianzas tiende a una matriz nula cuando $n \to \infty$,

$$\lim_{n\to\infty} V(\hat{\boldsymbol{\beta}}) = \lim_{n\to\infty} \frac{\sigma_u^2}{n}\left(\frac{\mathbf{X}'\mathbf{X}}{n}\right)^{-1} = \lim_{n\to\infty} \frac{\sigma_u^2}{n} \times \lim_{n\to\infty}\left(\frac{\mathbf{X}'\mathbf{X}}{n}\right)^{-1} = 0Q^{-1} = \mathbf{O}$$

□

La propiedad de consistencia significa que los estimadores de mínimos cuadrados tienden o convergen a los parámetros verdaderos al ir aumentando indefinidamente el tamaño de la muestra.

Observación 17. *El estimador de mínimos cuadrados se denomina ELIO para indicar que es un estimador lineal, insesgado y óptimo. El adjetivo óptimo indica que el estimador es el más eficiente o el de mínima varianza en la clase de estimadores lineales e insesgados.*

En resumen, el estimador de mínimos cuadrados $\hat{\boldsymbol{\beta}}$ cumple las propiedades estadísticas de linealidad, insesgadez, eficiencia y consistencia. Estas propiedades se consideran deseables y justifican el empleo del método de mínimos cuadrados como método de estimación en el marco del modelo lineal general clásico y nuestra preferencia por este método frente a otros métodos de estimación alternativos.

3.4. Propiedades estadísticas de $\hat{\sigma}_u^2$ y $\tilde{\sigma}_u^2$

PROPOSICIÓN 24. *La suma de cuadrados de los residuos* $\hat{\mathbf{u}}'\hat{\mathbf{u}}$ *es función cuadrática de las perturbaciones aleatorias,* $\hat{\mathbf{u}}'\hat{\mathbf{u}} = \mathbf{u}'\mathbf{M}\mathbf{u}$.

DEMOSTRACIÓN. Sabemos que $\hat{\mathbf{u}} = \mathbf{M}\mathbf{y}$ y $\mathbf{M}\mathbf{X} = \mathbf{0}$. Por tanto,

$$\hat{\mathbf{u}} = \mathbf{M}\mathbf{y} = \mathbf{M}\left[\mathbf{X}\boldsymbol{\beta} + \mathbf{u}\right] = \mathbf{M}\mathbf{u}$$

De aquí,

$$\hat{\mathbf{u}}'\hat{\mathbf{u}} = (\mathbf{M}\mathbf{u})'\mathbf{M}\mathbf{u} = \mathbf{u}'\mathbf{M}'\mathbf{M}\mathbf{u} = \mathbf{u}'\mathbf{M}\mathbf{u}$$

□

Vemos que la suma de cuadrados de los residuos es un estadístico, es decir, una función de las variables aleatorias $\{u_1, u_2, \ldots, u_n\}$. Su distribución de probabilidad puede, por tanto, derivarse de la distribución de probabilidad conjunta de las perturbaciones estocásticas $\{u_1, u_2, \ldots, u_n\}$.

TEOREMA 4. *La ratio* $\hat{\mathbf{u}}'\hat{\mathbf{u}}/\sigma_u^2$ *tiene una distribución Chi-cuadrado con* $n-k$ *grados de libertad, que se expresa sucintamente como*

$$\frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{\sigma_u^2} \sim \chi^2_{n-k}$$

DEMOSTRACIÓN. Usaremos los siguientes resultados sobre distribuciones de formas cuadráticas.

1. Sea $\mathbf{z} = (z_1 \; z_2 \; \ldots \; z_n)'$ un vector $n \times 1$ de variables aleatorias idéntica e independientemente distribuidas (*iid*) con distribución normal estándar, $\mathbf{z} \sim N(\mathbf{0}, \mathbf{I}_n)$. Entonces,

$$\mathbf{z}'\mathbf{z} = \sum_{i=1}^{n} z_i^2 \sim \chi^2_n$$

Demostración. Si $z_i \sim N(0,1)$, entonces $z_i^2 \equiv N(0,1)^2 \sim \chi^2_1$. Además, si $z_1, \ldots, z_n$ son variables aleatorias *iid* y si cada z_i tiene una distribución normal estándar, entonces la suma de los cuadradados $z_1^2 + \cdots + z_n^2$ tiene una distribución χ^2 con n grados de libertad.

2. Sea $\mathbf{u} = (u_1 \; u_2 \; \ldots \; u_n)'$ un vector $n \times 1$ de variables aleatorias idéntica e independientemente distribuidas como una normal con media 0 y varianza σ_u^2,

$\mathbf{u} \sim N(\mathbf{0}, \sigma_u^2 \mathbf{I}_n)$. Entonces,

$$\frac{1}{\sigma_u^2}\mathbf{u}'\mathbf{u} = \sum_{i=1}^{n} \left(\frac{u_i}{\sigma_u}\right)^2 \sim \chi_n^2$$

Demostración. Sea $\mathbf{z} \equiv \mathbf{u}/\sigma_u$. Entonces, $E(\mathbf{z}) = E(\mathbf{u}/\sigma_u) = 0$, $E(\mathbf{z}\mathbf{z}') = E(\mathbf{u}\mathbf{u}'/\sigma_u^2) = \mathbf{I}_n$, y $\mathbf{z} \sim N(0, \mathbf{I}_n)$. Por el resultado 1, $\mathbf{z}'\mathbf{z} \equiv \mathbf{u}'\mathbf{u}/\sigma_u^2 \sim \chi_n^2$.

3. Sea $\mathbf{u} \sim N(\mathbf{0}, \sigma_u^2 \mathbf{I}_n)$ y sea $\mathbf{M}$ una matriz simétrica e idempotente de rango $n-k$. Entonces

$$\frac{1}{\sigma_u^2}\mathbf{u}'\mathbf{M}\mathbf{u} \sim \chi_{n-k}^2$$

Demostración. Sean $\mathbf{P}$ y $\mathbf{\Lambda}$ las matrices de autovectores y autovalores de $\mathbf{M}$, $\mathbf{M}\mathbf{P} = \mathbf{P}\mathbf{\Lambda}$. Por ser $\mathbf{M}$ simétrica, $\exists \mathbf{P}^{-1} = \mathbf{P}'$ y $\mathbf{M} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}'$. Por ser $\mathbf{M}$ idempotente, $\mathbf{M} = \mathbf{P}\mathbf{\Lambda}^2\mathbf{P}'$, los autovalores tienen que ser iguales a 1 ó 0. Como $tr\mathbf{M} = tr\mathbf{\Lambda} = n-k$ se deduce que de los n autovalores, $n-k$ son iguales a uno y k son iguales a cero. Define $\mathbf{u}_* = \dfrac{1}{\sigma_u}\mathbf{P}\mathbf{u}$. Entonces, $\mathbf{u}_* \sim N(\mathbf{0}, \mathbf{I}_n)$ porque $\mathbf{P}'\mathbf{P} = \mathbf{I}_n$. Luego

$$\frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{\sigma_u^2} = \frac{1}{\sigma_u^2}\mathbf{u}'\mathbf{M}\mathbf{u} = \frac{1}{\sigma_u^2}\mathbf{u}'\mathbf{P}'\mathbf{\Lambda}\mathbf{P}\mathbf{u} = \mathbf{u}_*'\mathbf{\Lambda}\mathbf{u}_* = \sum_{i=1}^{n-k} u_{*i}^2 \sim \chi_{n-k}^2$$

□

Proposición 25. *$\hat{\sigma}_u^2 = \hat{\mathbf{u}}'\hat{\mathbf{u}}/(n-k)$ es un estimador insesgado de σ_u^2 con varianza $2\sigma_u^4/(n-k)$.*

Demostración. La esperanza matemática de una variable aleatoria z con distribución *Chi*-cuadrado con m grados de libertad es igual a los grados de libertad m, $E(z) = m$. Por tanto,

$$E\left(\frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{\sigma_u^2}\right) = (n-k)$$

De aquí, $E(\hat{\mathbf{u}}'\hat{\mathbf{u}}) = (n-k)\sigma_u^2$ y

$$E(\hat{\sigma}_u^2) = E\left(\frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{n-k}\right) = \sigma_u^2$$

La varianza de $z \sim \chi_m^2$ es igual a dos veces los grados de libertad, $var(z) = 2m$. Por tanto,

$$var\left(\frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{\sigma_u^2}\right) = 2(n-k)$$

De aquí, $var(\hat{\mathbf{u}}'\hat{\mathbf{u}}) = 2(n-k)\sigma_u^4$ y

$$var(\hat{\sigma}_u^2) = \frac{var(\hat{\mathbf{u}}'\hat{\mathbf{u}})}{(n-k)^2} = \frac{2\sigma_u^4}{n-k}$$

□

Observación 18. *La esperanza matemática de la suma de cuadrados de los residuos puede obtenerse sin conocer su distribución de probabilidad*

$$\begin{aligned}
E(\hat{\mathbf{u}}'\hat{\mathbf{u}}) =& E(\mathbf{u}'\mathbf{M}\mathbf{u}) && \textit{Proposición 24}\\
=& E(tr\mathbf{u}'\mathbf{M}\mathbf{u}) && \textit{Propiedad: tr(escalar) = escalar}\\
=& E(tr\mathbf{M}\mathbf{u}\mathbf{u}') && \textit{Propiedad: } tr(\mathbf{ABC}) = tr(\mathbf{CBA})\\
=& trE(\mathbf{M}\mathbf{u}\mathbf{u}') && \textit{Propiedad: } E(\sum_{i=1}^{n} z_i) = \sum_{i=1}^{n} E(z_i)\\
=& tr\left[\mathbf{M}E(\mathbf{u}\mathbf{u}')\right] && \textit{Supuesto: } \mathbf{X} \textit{ es una matriz fija}\\
=& tr\left[\mathbf{M}(\sigma_u^2\mathbf{I}_n)\right] = tr\left[\sigma_u^2\mathbf{M}\right] && \textit{Supuesto: } E(\mathbf{u}\mathbf{u}') = \sigma_u^2\mathbf{I}_n\\
=& \sigma_u^2 tr\mathbf{M} && \textit{Propiedad: factor común}\\
=& \sigma_u^2(n-k) && \textit{Propiedad: } tr\mathbf{M} = (n-k)
\end{aligned}$$

COROLARIO 8. $\tilde{\sigma}_u^2 = \hat{\mathbf{u}}'\hat{\mathbf{u}}/n$ *es un estimador sesgado de* σ_u^2*, siendo el sesgo* $B(\tilde{\sigma}_u^2) = (-k/n)\sigma_u^2$.

DEMOSTRACIÓN. De la relación entre $\hat{\sigma}_u^2$ y $\tilde{\sigma}_u^2$

$$\tilde{\sigma}_u^2 = \frac{n-k}{n}\hat{\sigma}_u^2$$

se tiene que $E(\tilde{\sigma}_u^2) = \sigma_u^2 - (k/n)\sigma_u^2$. □

PROPOSICIÓN 26. $\tilde{\sigma}_u^2 = \hat{\mathbf{u}}'\hat{\mathbf{u}}/n$ *es un estimador consistente de* σ_u^2.

DEMOSTRACIÓN. El estimador $\tilde{\sigma}_u^2$ converge en media cuadratica al verdadero parámetro σ_u^2

1. $\lim_{n\to\infty} B(\tilde{\sigma}_u^2) = \lim_{n\to\infty}(-k/n)\sigma_u^2 = 0$
2. $\lim_{n\to\infty} var(\tilde{\sigma}_u^2) = \lim_{n\to\infty} \dfrac{2(n-k)}{n^2}\sigma_u^4 = 0$

□

Observación 19. *Mientras que el estimador* $\hat{\boldsymbol{\beta}}$ *resulta de un proceso de minimización, el estimador* $\hat{\sigma}_u^2$ *se construye para que sea insesgado.*

3.5. Resumen

1. Un estimador es insesgado si su valor esperado coincide con el parámetro que se desea estimar.
2. Un estimador es consistente si la estimación del parámetro en muestras grandes es el parámetro que se desea estimar.
3. Un estimador es eficiente dentro de una clase de estimadores si su varianza es menor que la de los otros estimadores.
4. Bajo los supuestos básicos, el estimador de mínimos cuadrados es ELIO (en inglés, BLUE: Best Linear Unbiased Estimator).
5. Bajo el supuesto de normalidad de las perturbaciones, el estimador de mínimos cuadrados tiene una distribución normal multivariante.
6. El error estándar de la regresión es la raíz cuadrada de la varianza muestral de los residuos.

7. La precisión de los estimadores es inversamente proporcional al error estándar de la regresión.

Palabras clave

Modelo clásico de regresión	Regresores no estocásticos
Distribución normal multivariante	Multicolinealidad
Vector de medias	Homocedasticidad
Matriz de varianzas y covarianzas	Correlación serial

3.6. Ejercicios

1. Use el proceso generador de datos

$$Y_t = 1{,}0 + 0{,}5t + u_t \quad u_t \sim N(0,1)$$

para generar 10 muestras de 25 observaciones $(Y_1, \ldots, Y_{25})$. Utilice cada muestra para estimar la regresión lineal simple de Y_t sobre la tendencia lineal t. Compare las estimaciones de β_1 y β_2 obtenidas en cada muestra con los valores verdaderos. Calcule la media y desviación típica de las 10 estimaciones de β_1 y β_2, ¿qué puede decir sobre la propiedad de insesgadez?. Genere después una muestra de 200 observaciones, y estime la regresión simple: ¿que puede decir sobre la propiedad de consistencia?.

2. Discuta las siguientes proposiciones:
   a) El supuesto $\rho(\mathbf{X}) = k$ implica que las variables explicativas son ortogonales.
   b) Si para estimar la ecuación de regresión simple, $y_i = \beta_1 + \beta_2 X_i + u_i$, sólo se dispone de un dato, $i = 1$, entonces el estimador de mínimos cuadros de los parámetros está indeterminado.
   c) Los momentos respecto al origen de la perturbación aleatoria u_i coinciden con sus momentos centrados.
   d) El estimador de la varianza residual es un estimador lineal.

3. Demuestre que $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}$. Derive la distribución de probabilidad del estimador $\hat{\boldsymbol{\beta}}$ a partir de la distribución de probabilidad de $\mathbf{u}$.

4. Demuestre que la submatriz de covarianzas de $(\hat{\beta}_i, \hat{\beta}_j)$ es semidefinida positiva. Utilice este resultado para demostrar que

$$cov(\hat{\beta}_i, \hat{\beta}_j)^2 \leq var(\hat{\beta}_i)var(\hat{\beta}_j)$$

¿Qué puede decir sobre la correlación entre $\hat{\beta}_i$ y $\hat{\beta}_j$?

5. Demuestre que $Var(\hat{y}_i)$ puede escribirse como

$$Var(\hat{y}_i) = \sum_{j=1}^{k} x_{ji}^2 Var(\hat{\beta}_j) + 2\sum_{j=2}^{k}\sum_{h=1}^{j-1} x_{ji}x_{ki}cov(\hat{\beta}_j, \hat{\beta}_h)$$

6. Demuestre que

$$E\left[(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})\right] = \left[(E\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'(E\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})\right] + E\left[(\hat{\boldsymbol{\beta}} - E\hat{\boldsymbol{\beta}})'(\hat{\boldsymbol{\beta}} - E\hat{\boldsymbol{\beta}})\right]$$
$$= \sum_{i=1}^{k} sesgo^2(\hat{\beta}_i) + \sum_{i=1}^{k} var(\hat{\beta}_i)$$

7. Derive las propiedades estadísticas de los residuos mínimo-cuadráticos, $E(\hat{\mathbf{u}})$ y $V(\hat{\mathbf{u}})$.
8. Demuestre que $V(\hat{u}_t) = (1 - h_t)\sigma_u^2$, en donde $h_t = \mathbf{x}_t'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_t$.