Errores estándar

Econometría Aplicada y Ciencia de Datos - CIDE Fall 2026

Carlos Brito

Introducción

- Estimar bien los coeficientes no es suficiente: también necesitamos saber qué tan precisas son esas estimaciones

$Var(u|X) = \sigma^2$

- Estándar de MCO: las cuales asumen errores **homocedásticos** e **independientes**

↳ $cov(u_i, u_j) = 0$ for all $i \neq j$

- Estos supuestos rara vez se sostienen con datos reales. Dos problemas recurrentes:
  - **Heterocedasticidad**: la varianza del error no es constante
  - **Agrupamiento**: observaciones correlacionadas dentro de grupos (salones, hogares, entidades, etc.)

- Ignorar estos problemas no sesga los coeficientes de MCO, pero sí distorsiona los errores estándar (típicamente subestimándolos)

↳ $E[\hat{\beta}_{OLS}] = \beta$ { main assumption: $E[u|X] = 0$

$$\hat{\beta}_{OLS} = (X'X)^{-1}(X'y) \xrightarrow{y = X'\beta + \mu} \hat{\beta}_{OLS} = \beta + (X'X)^{-1}X'\mu$$

Errores estándar robustos

$$V(\hat{\beta}_{OLS}) = \underbrace{(X'X)^{-1}X'\Omega X(X'X)^{-1}}_{\Omega = \sigma^2 I \text{ (homoc)} \;\Rightarrow\; \sigma^2(X'X)^{-1}}; \quad \Omega = E[\mu\mu']$$

Errores estándar robustos

- Recordemos que con errores homocedásticos, la matriz de varianzas del estimador de MCO puede ser estimada como:

$$\Omega = \sigma^2 I$$

$$\hat{V}(\hat{\beta}_{OLS}^{H}) = \hat{\sigma}^2 (X'X)^{-1}$$

$$\hat{\sigma}^2 = \frac{1}{n-K} \hat{u}_i^2 \quad ; \quad \hat{u}_i^2 = (y - X_i' \hat{\beta}_{OLS})^2$$

- Una primera *desviación* respecto a los errores clásicos ocurre cuando relajamos el supuesto de homocedasticidad
- La varianza asintótica robusta a heterocedasticidad

$$\hat{V}(\hat{\beta}_{OLS}^{R}) = (X'X)^{-1} X' \underbrace{\Omega}_{\hat{u}\hat{u}'} X (X'X)^{-1}$$

Errores robustos a la heterocedasticidad

- Un estimador de la varianza del estimador de MCO que no asume homocedasticidad es el estimador propuesto por White (1980)

$$X'\mu\mu'X = (X'\mu)(\mu'X) = (X'\mu)(X'\mu)' = \left(\sum_i X_i \mu_i\right)\left(\sum_j X_j \mu_j\right)'$$

$$= \sum_i \sum_j X_i X_j' \mu_i \mu_j \xrightarrow{\text{by indep.}} = \sum_i \mu_i^2 X_i X_i'$$

$\mu_i \mu_j$ is only $\neq$ zero for $i=j$

$$\hat{V}\left(\hat{\beta}^R_{OLS}\right) = (X'X)^{-1}\left(\sum_i \mu_i^2 X_i X_i'\right)(X'X)^{-1}$$

Errores estándar robustos

Consideremos la *carnita* del sándwich

$$\sum_i \hat{u}_i^2 x_i x_i \equiv \sum_i \hat{\psi}_i x_i x_i'$$

- Dependiendo de cómo se especifique $\hat{\psi}_i$, obtenemos distintas versiones del estimador de varianzas robusto:

1 La propuesta de White original es:

$$HC0: \quad \hat{\psi}_i = \hat{u}_i^2$$

Este estimador es asintóticamente consistente

2 En muestras pequeñas, muchas veces se emplea la siguiente corrección:

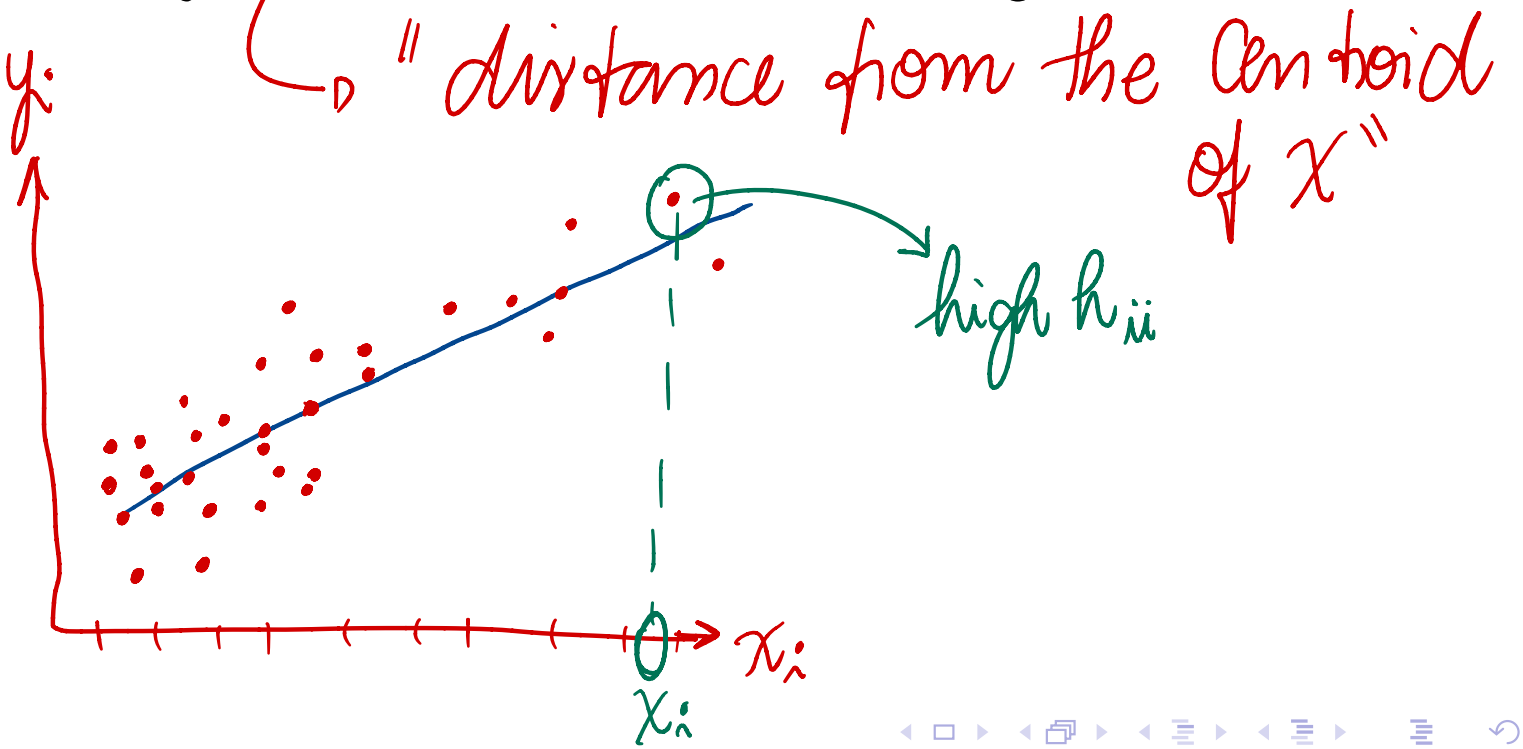
$$HC1: \quad \hat{\psi}_i = \frac{N}{N-k} \hat{u}_i^2$$

Desviación a la influencia

- Un par de resultados nos ayudarán a entender qué hacen las otras correcciones a la matriz robusta en el software
- Definimos la **influencia** de la observación i como:

$$h_{ii} = X_i'(X'X)^{-1}X_i$$

h_{ii} nos dice qué tanto *jala* la observación i a la línea de regresión



Desviación a la influencia

- Un par de resultados nos ayudarán a entender qué hacen las otras correcciones a la matriz robusta en el software
- Definimos la **influencia** de la observación i como:

$$h_{ii} = X_i'(X'X)^{-1}X_i$$

h_{ii} nos dice qué tanto *jala* la observación i a la línea de regresión

- En una regresión con un solo regresor x, se puede mostrar que la influencia de la observación i es:

$$h_{ii} = \frac{1}{N} + \frac{(x_i - \bar{x})^2}{\sum(x_j - \bar{x})^2}$$

- es decir, que la influencia se incrementa cuando x_i se aleja de la media
- La influencia es un número entre 0 y 1 y además $\sum_i h_{ii} = k$, siendo k el número de regresores

Errores estándar robustos

- Algunos autores sugieren usar la influencia en la matriz de varianzas robusta
- Se proponen algunas alternativas:

$$HC2: \quad \hat{\psi}_i = \frac{1}{1 - h_{ii}} \hat{u}_i^2$$

$$HC3: \quad \hat{\psi}_i = \frac{1}{(1 - h_{ii})^2} \hat{u}_i^2$$

- Long & Ervin (2000) realizaron un experimento de simulación y recomendaron usar *HC3 en muestras pequeñas*, por lo que el paquete *sandwich* en R usa *HC3* por default
- Es importante tener en cuenta qué tipo de errores estándar piden que el software calcule

Errores agrupados

Errores agrupados

Surgen naturalmente cuando las observaciones están agrupadas

- Niños en salones de clase
- Hogares en localidades
- Solicitudes de empleo en una empresa
- Ahorradoras en un banco

Errores agrupados

Surgen naturalmente cuando las observaciones están agrupadas

- Niños en salones de clase
- Hogares en localidades
- Solicitudes de empleo en una empresa
- Ahorradoras en un banco

El supuesto de errores independientes claramente no se cumple
Pensemos en un problema simple para entender la intuición:

$$y_{ig} = \beta_0 + \beta_1 x_g + e_{ig}$$

Aquí, x_g es un regresor que es el mismo para todos los miembros del grupo g y asumamos que todos los grupos tienen tamaño n

Errores agrupados

- Podemos mostrar que la correlación de errores entre dos observaciones i y j que pertenecen a g es

$$\underbrace{\rho_e}_{\substack{\text{coeficiente de correlación} \\ \text{intraclase residual}}} \underbrace{\sigma_e^2}_{\substack{\text{varianza} \\ \text{residual}}}$$

Errores agrupados

- Podemos mostrar que la correlación de errores entre dos observaciones i y j que pertenecen a g es

$$E(e_{ig}e_{jg}) = \overbrace{\rho_e}^{\substack{\text{coeficiente de correlación} \\ \text{intraclase residual}}} \underbrace{\sigma_e^2}_{\text{varianza residual}}$$

- Le damos una estructura aditiva a los errores:

$$e_{ig} = \nu_g + \eta_{ig}$$

  - donde ν_g captura toda la correlación dentro del grupo
  - η_{ig} es un error idiosincrático con media cero e independiente de cualquier otro η_{jg}

- Como queremos analizar el problema del agrupamiento, asumimos que tanto ν_g y η_{ig} son homocedásticos

Errores agrupados

$\rho_e = \frac{cov(e)}{Var\,e} = \frac{\sigma_v^2}{\sigma_v^2 + \sigma_\eta^2}$

- Con esta estructura de errores, el coeficiente de correlación intraclase es:

$E[e_{ig} e_{jg}] = E[(v_g + \eta_{ig})(v_g + \eta_{jg})] = E[v_g^2 + v_g \eta_{jg} + v_g \eta_{ig} + \eta_{ig}\eta_{jg}] = \sigma_v^2$

$= \rho_e \cdot (\sigma_v^2 + \sigma_\eta^2)$

$\rho_e \to \sigma_v^2/(\sigma_v^2 + \sigma_\eta^2)$

- Deberíamos calcular la matriz de varianzas $V_C(\hat{\beta})$ tomando en cuenta esta estructura

Errores agrupados

- Con esta estructura de errores, el coeficiente de correlación intraclase es:

$$\rho_e = \frac{\sigma^2_v}{\sigma^2_v + \sigma^2_\eta} = \frac{cov(e)}{Var(e)}$$

- Deberíamos calcular la matriz de varianzas $V_C(\hat{\beta})$ tomando en cuenta esta estructura
- ¿Qué pasa si hacemos MCO en el contexto de este problema?
  - Moulton (1984) muestra que:

$$\frac{V_C(\hat{\beta})}{V_{MCO}(\hat{\beta})} = 1 + (n-1)\rho_e$$

  - A $\sqrt{\frac{V_C(\hat{\beta})}{V_{MCO}(\hat{\beta})}}$ se le conoce como el *factor de Moulton*

Factor de Moulton

- Factor de Moulton: qué tanto sobreestimamos la precisión al ignorar la correlación intra clase

Visto de otro modo:

$$V_C(\hat{\beta}) = (1 + (n-1)\rho_e) V_{OLS}(\hat{\beta})$$

- $\uparrow \rho_e$: más deberíamos *inflar* los errores de MCO.
- El factor aumenta con menos conglomerados/clases: hay menos información independiente en la muestra.

Factor de Moulton

- Factor de Moulton: qué tanto sobreestimamos la precisión al ignorar la correlación intra clase

Visto de otro modo:

$$V_C(\hat{\beta}) = (1 + (n-1)\rho_e) V_{OLS}(\hat{\beta})$$

  - $\uparrow \rho_e$: más deberíamos *inflar* los errores de MCO.
  - El factor aumenta con menos conglomerados/clases: hay menos información independiente en la muestra.

$\rho_e = 1$: todas las y_{ig} dentro del mismo g son iguales.

- Entonces el factor de Moulton es simplemente $\sqrt{n}$.
- Matriz de varianzas correcta se obtendría multiplicando por n la matriz $V_{MCO}(\hat{\beta})$

$$V_C(\hat{\beta}) = nV_{MCO}(\hat{\beta})$$

Errores agrupados en general

En general, x_{ig} varía a nivel individual y tenemos grupos de tamaño n_g
En este caso, el factor de Moulton es la raíz cuadrada de:

$$\frac{V_C(\hat{\beta})}{V_{MCO}(\hat{\beta})} = 1 + \left(\frac{V(n_g)}{\bar{n}} + \bar{n} - 1 \right) \rho_x \rho_e$$

donde $\bar{n}$ es el tamaño promedio del grupo y ρ_x es la correlación intraclase de x_{ig}

- No es necesario asumir una forma para ρ_x (se puede calcular)
- Noten que el error que cometemos es más grande entre más heterogéneo es el tamaño de grupos y entre más grande es ρ_x
- Por tanto, cuando el tratamiento no varía entre grupos, este error es grande: la agrupación nos preocupa especialmente cuando el regresor de interés (la condición de tratamiento) es constante dentro de los grupos.

$\rho_x = 1 \Rightarrow x_{ig} = x_g$

Soluciones para errores agrupados

- Solución paramétrica: calcular directamente el factor de Moulton e inflar los errores de MCO
- Bootstrap por bloques: ver más adelante el concepto de bootstrap
- Estimar los errores agrupados (*clustered standard errors*)

Errores estándar agrupados

Con errores agrupados podemos escribir el estimador de MCO como

$$\hat{\beta} = \beta + (X'X)^{-1}X'u$$

$$(X'X)^{-1}\left(\sum_{g=1}^{G} X_g u_g\right)$$

Suponiendo independencia entre g y correlación dentro de cada grupo:

$$E(u_{ig}u_{jg'}|x_{ig}x_{jg'}) = 0$$

excepto cuando $g = g'$

En este caso, el estimador de MCO tiene una varianza asintótica dada por

$$V(\hat{\beta}^{R}_{OLS}) = (X'X)^{-1}\left(\sum_i \hat{u}_i^2 X_i X_i'\right)(X'X)^{-1}$$

$$V_C(\hat{\beta}_{MCO}) = (X'X)^{-1}\left(\sum_{g=1}^{G} X_g' u_g u_g' X\right)(X'X)^{-1}$$

Errores estándar agrupados

Con errores heterocedásticos, pero sin agrupamiento, la matriz de varianzas de White (1980) tiene una estructura como sigue:

$$\hat{V}(\hat{\beta}_R) = (X'X)^{-1}X'\hat{\Sigma}X(X'X)^{-1}$$

Donde

$$\hat{\Sigma} = \begin{pmatrix} \hat{u}_1^2 & 0 & 0 & \dots & 0 \\ 0 & \hat{u}_2^2 & 0 & \dots & 0 \\ \vdots & & & & \\ 0 & & & \dots & \hat{u}_n^2 \end{pmatrix}$$

Errores estándar agrupados

- Para estimar la varianza con errores agrupados empleamos una generalización de la propuesta de White para errores robustos
- Si $G \to \infty$, el estimador de la matriz de errores agrupados robusta (CRVE) es consistente para estimar $V(\hat{\beta})$:

$$\hat{V}_{CR}(\hat{\beta}) = (X'X)^{-1} \left(\sum_{g=1}^{G} X_g' \hat{u}_g \hat{u}_g' X_g \right) (X'X)^{-1}$$

donde $\hat{u}_g \hat{u}_g'$ es la matriz de varianzas para los individuos del grupo g
De manera compacta

$$\hat{V}_{CR}(\hat{\beta}) = (X'X)^{-1} X' \hat{\Sigma} X (X'X)^{-1}$$

Errores estándar agrupados

Y en este caso la matriz $\hat{\Sigma}$ tiene una estructura agrupada:

$$\hat{\Sigma} = \begin{pmatrix}
\hat{u}_{1,1}^2 & \hat{u}_{1,1}\hat{u}_{2,1} & \dots & \hat{u}_{1,1}\hat{u}_{n,1} & 0 & 0 & \dots & 0 & \dots & 0 & 0 & \dots & 0 \\
\hat{u}_{2,1}\hat{u}_{1,1} & \hat{u}_{2,1}^2 & \dots & \hat{u}_{2,1}\hat{u}_{n,1} & 0 & 0 & \dots & 0 & \dots & 0 & 0 & \dots & 0 \\
\vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & & \vdots & \vdots & & \vdots \\
\hat{u}_{n,1}\hat{u}_{1,1} & \hat{u}_{n,1}\hat{u}_{2,1} & \dots & \hat{u}_{n,1}^2 & 0 & 0 & \dots & 0 & \dots & 0 & 0 & \dots & 0 \\
0 & 0 & \dots & 0 & \hat{u}_{1,2}^2 & \hat{u}_{1,2}\hat{u}_{2,2} & \dots & \hat{u}_{1,2}\hat{u}_{n,2} & \dots & 0 & 0 & \dots & 0 \\
0 & 0 & \dots & 0 & \hat{u}_{2,2}\hat{u}_{1,2} & \hat{u}_{2,2}^2 & \dots & \hat{u}_{2,2}\hat{u}_{n,2} & \dots & 0 & 0 & \dots & 0 \\
\vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & & \vdots & \vdots & & \vdots \\
0 & 0 & \dots & 0 & \hat{u}_{n,2}\hat{u}_{1,2} & \hat{u}_{n,2}\hat{u}_{2,2} & \dots & \hat{u}_{n,2}^2 & \dots & 0 & 0 & \dots & 0 \\
\vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & & \vdots & \vdots & & \vdots \\
0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & \dots & \hat{u}_{1,G}^2 & \hat{u}_{12,G}\hat{u}_{2,G} & \dots & \hat{u}_{1,G}\hat{u}_{n,G} \\
0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & \dots & \hat{u}_{2,G}\hat{u}_{1,G} & \hat{u}_{2,G}^2 & \dots & \hat{u}_{2,G}\hat{u}_{n,G} \\
\vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & & \vdots & \vdots & & \vdots \\
0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & \dots & \hat{u}_{n,G}\hat{u}_{1,G} & \hat{u}_{n,G}\hat{u}_{2,G} & \dots & \hat{u}_{n,G}^2
\end{pmatrix}$$

Errores estándar agrupados

- El resultado asintótico de consistencia depende de que $G \to \infty$
- Si G está fijo, no importa qué tan grande sea N, $\hat{V}_{CRVE}(\hat{\beta})$ no será consistente
- Algunos paquetes ajustan esta matriz de varianzas haciendo una corrección parecida a $HC1$, pero ahora tomando en cuenta también G y no solo N (ver por ejemplo, *vcovCR* en R)
- Con pocos grupos, subestimamos los errores estándar y rechazamos la H_0 más veces de lo que deberíamos (*over-rejection*)
- Si tenemos pocos grupos, recurrimos a otras soluciones (ver Cameron y Miller, 2015)
  - Inflar los errores con un corrector de sesgo
  - Bootstrap agrupado con refinamiento asintótico
- La recomendación práctica es que se tomen en serio el problema de los pocos clusters
- ¿Cuánto es poco? Cameron y Miller (2015) citan 50.