Notas de Econometría

Pavel Solís
2025

6 Análisis de Regresión Múltiple: Cuestiones Prácticas

Temas importantes para aplicar RM en la práctica

6.1 Efecto del cambio de unidades de medición sobre estadísticos MCO

Las magnitudes de $\widehat{\beta}_j$ se pueden cambiar al cambiar las unidades de medida de x_j

> Ejemplo.
> regress pesolb alturapg edadmss
> regress pesokg alturapg edadmss
> regress pesolb alturapg edadans
> regress pesokg alturacm edadans

- Cambios porcentuales y elasticidades son invariantes a las unidades de medición de y o x_j
  - Cambiar las unidades de medición de la variable dependiente cuando aparece en logaritmo, no afecta a los estimados de las pendientes

    $\log(c_1 y_i) = \log(c_1) + \log(y_i)\, \forall\, c_1 > 0$, y el nuevo intercepto será $\log(c_1) + \widehat{\beta}_0$
  - Cambiar las unidades de medición de cualquier variable independiente cuando aparece en logaritmo, solo afecta al intercepto

¿Cómo cambian los estadísticos t y F?

- Efectos de reescalar las variables dependiente e independientes:
  - Coeficientes, errores estándar, estadísticos t y F, intervalos de confianza cambian de forma que se preservan los efectos identificados y los resultados de las pruebas
- Llegamos a las mismas conclusiones sin importar cómo se midan las variables
- Razones (cosméticas):
  - Reducir el número de ceros después del punto decimal
  - Mejorar la apariencia de ecuaciones estimadas
  - Ej. Cantidades en pesos

6.1.1 Coeficientes beta

A veces una variable clave esta medida en una escala difícil de interpretar (ej. IQ)

- En esos casos, podemos preguntarnos el efecto cuando la variable es una desviación estándar más alta

Podemos obtener regresiones cuando <u>todas</u> las variables están estandarizadas

- *z-score*: A cada observación se le resta $\bar{x}$ y se divide por su desviación estándar

$\widehat{b}_j$ son los coeficientes estandarizados o coeficientes beta

$$\widehat{b}_j = \frac{\widehat{\sigma}_j}{\widehat{\sigma}_y}\widehat{\beta}_j \qquad \forall\ j = 1, 2, \ldots, k$$

- Medimos efectos en desviaciones estándar no en unidades originales de y y x_j
- Si x_j aumenta en 1 desviación estándar, $\widehat{y}$ cambia en $\widehat{b}_j$ desviaciones estándar
- $\widehat{b}_j$ de las variables independientes se pueden comparar
- Ej. Desempeño de un estudiante relativo al resto de la clase
- En Stata, regress y x_1 x_2 x_3, beta

6.2 Formas funcionales

Podemos emplear lo siguiente de forma aislada o conjunta

- Logaritmos
- Términos cuadráticos
- Interacciones

6.2.1 Formas funcionales logarítmicas

Ejemplo. Salario y estudios universitarios

$$\log(precio) = \beta_0 + \beta_1 \log(nox) + \beta_2 habit + u$$

$$\widehat{\log(precio)} = \underset{(0.19)}{9.73} - \underset{(0.066)}{0.718} \log(nox) + \underset{(0.019)}{0.306}\, habit$$

$$n = 506,\ \mathrm{R}^2 = 0.514$$

- ¿Cuánto baja el precio si *nox* aumenta 1%, fijando *habit*?
- ¿Cuánto sube el precio si *habit* aumenta en 1, fijando *nox*?
- La aproximación $\%\Delta x \approx 100 \times \Delta \log(x)$ aplica para cambios pequeños en x
  - Aproximación imprecisa cuando los cambios en x son grandes

Cálculo del cambio porcentual exacto

$$\%\Delta\widehat{y} = 100\left[\exp\left(\widehat{\beta}_2\Delta x_2\right) - 1\right]$$

- Recordatorio: $\%\Delta x = 100\left(\frac{\Delta x}{x_0}\right) = 100\left(\frac{x_1 - x_0}{x_0}\right)$
- Derivación:

$$\widehat{\log(y)} = \widehat{\beta}_0 + \widehat{\beta}_1 \log(x_1) + \widehat{\beta}_2 x_2 \implies \Delta\widehat{\log(y)} = \widehat{\beta}_2 \Delta x_2 \text{ fijando } x_1$$

$$\Delta \log\left(\frac{\widehat{y}_1}{\widehat{y}_0}\right) = \widehat{\beta}_2\Delta x_2 \implies \frac{\widehat{y}_1}{\widehat{y}_0} = \exp\left(\widehat{\beta}_2\Delta x_2\right) \implies \frac{\widehat{y}_1}{\widehat{y}_0} - 1 = \exp\left(\widehat{\beta}_2\Delta x_2\right) - 1$$

Ejemplo (continuación)

- Si $\Delta x_2 = 1$,

$$\%\Delta\widehat{precio} = 100\left[\exp\left(0.306\right) - 1\right] = 35.8\%$$

- Si $\Delta x_2 = -1$,

$$\%\Delta\widehat{precio} = 100\left[\exp\left(-0.306\right) - 1\right] = -26.4\%$$

- $|-26.4\%| <$ aproximación $< |35.8\%|$

- Ajuste no es crucial para cambios pequeños

Ejercicio. Cálcula la aproximación y el cambio porcentual exacto si $\widehat{\beta}_{stratio} = -0.052$ y $\Delta stratio = 1$ o $\Delta stratio = 5$

- $\log(x_j)$ captura rendimientos decrecientes de x_j sobre y

Razones para usar logaritmos:

- Interpretación útil de los coeficientes
- Coeficientes de pendiente invariantes ante reescalamiento
  - Ignoramos unidades de medición
- Cuando $y > 0$, variable dependiente $\log(y)$ satisface supuestos MLC más fácil que variable dependiente y
- Cuando $x > 0$, la distribución condicional tiende a ser heterocedástica o sesgada, $\log(y)$ mitiga ambas
- $\log(z)$ reduce rango de z (ventas, salario, población) y hace a MCO menos sensible a valores extremos
  - Cuidado: $\log(z)$ puede crear valores extremos si $0 < z < 1$ (es una proporción)

Usos comunes:

- En logaritmos: dinero, salario, ventas, población, empleados, alumnos
- En niveles:
  - Tiempo (educación, experiencia, antigüedad, edad)
  - Proporciones y porcentajes
    - Para interpretar cambio en puntos porcentuales
    - Cambios porcentuales con logaritmos

Limitante: Logaritmos no se puede usar si variable toma valores negativos o cero

- Cuando $z \geq 0$ con pocos 0's, podemos usar $\log(1+z)$ con interpretación $\log(z)$
  - Considera que $\%\Delta z$ no está definido cuando $z_0 = 0$
  - Si hay muchos 0's, se pierden observaciones

Consideraciones:

- Si variable dependiente es $\log(y)$, usar función exponencial para predecir y
- R^2 no se pueden comparar entre y y $\log(y)$ porque la variación es diferente

6.2.2 Modelos con términos cuadráticos

Funciones cuadráticas se ocupan para capturar efectos marginales

- Decrecientes
- Crecientes

Si $y = \beta_0 + \beta_1 x + \beta_2 x^2 + u$,

$$\widehat{y} = \widehat{\beta}_0 + \widehat{\beta}_1 x + \widehat{\beta}_2 x^2 + u \implies \Delta\widehat{y} \approx (\widehat{\beta}_1 + 2\widehat{\beta}_2 x)\Delta x \implies \frac{\Delta\widehat{y}}{\Delta x} \approx \widehat{\beta}_1 + 2\widehat{\beta}_2 x$$

- Pendiente estimada de la relación entre x y y (depende del valor de x): $\widehat{\beta}_1 + 2\widehat{\beta}_2 x$
- Resume el efecto de x sobre y
- En lugar de x también podemos sustituir media, mediana, cuartiles

El punto óptimo es:

$$x^* = \frac{-\widehat{\beta}_1}{2\widehat{\beta}_2}$$

- Es positivo si $\widehat{\beta}_1$ y $\widehat{\beta}_2$ tienen signos opuestos
- Es negativo si $\widehat{\beta}_1$ y $\widehat{\beta}_2$ tienen el mismo signo
- Si $\widehat{\beta}_2 > 0$, efecto creciente en y y x^* es mínimo
- Si $\widehat{\beta}_2 < 0$, efecto decreciente en y y x^* es máximo
- Si $x > 0$ y

- $\widehat{\beta}_2 > 0$, $\min \mathbb{E}(y)$ cuando $x = 0$, aumentos en x tienen un efecto positivo y creciente en y
- $\widehat{\beta}_2 < 0$, $\max \mathbb{E}(y)$ cuando $x = 0$, aumentos en x tienen un efecto negativo y decreciente en y

Con datos, se puede calcular x^* y ver si tiene sentido

> Ejemplo. Salario y experiencia
>
> $$\begin{array}{lllll} \widehat{salario} = & 3.73 & + \ 0.298\, exper & - & 0.0061\, exper^2 \\ & (0.35) & \ \ (0.091) & & (0.0009) \end{array}$$
>
> $$n = 526, \ \mathrm{R}^2 = 0.093$$
>
> - ¿x^*?
> - Cuando $x < x^*$, aumentos en x tienen un efecto positivo en y
> - Cuando $x = x^*$, el rendimiento de la experiencia es 0
> - Cuando $x < x^*$, aumentos en x tienen un efecto negativo en y

Significado de x^*:

- Pocas personas en la muestra con $exper > 24$, entonces parte $x > x*$ se puede ignorar
- Tal vez rendimiento de *exper* sí es negativo en algún punto, pero no creíble que sea cuando $exper = 24$
- El efecto de *exper* sobre el salario está sesgado
  - Faltan controles
  - Forma funcional no es correcta
  - Ej. $\log(salario)$ $exper$ $exper^2$ $educ$ puede dar $x^* \approx 29$

> Ejemplo. Efecto del número de habitaciones en el precio de casas
>
> $$\begin{array}{llllll} \widehat{\log(precio)} = & 13.39 & - & 0.902 \log(nox) & - & 0.87 \log(dist) \\ & (0.57) & & (0.115) & & (0.043) \\ & -\,0.545\, habit & + & 0.062\, habit^2 & - & 0.48\, stratio \\ & (0.165) & & (0.013) & & (0.006) \end{array}$$
>
> $$n = 506, \qquad \mathrm{R}^2 = 0.603$$
>
> - ¿t_{habit} es significativo?
> - ¿$habit^*$?

- $x < x^*$ en solo 1% de observaciones, podemos ignorar esa parte
- Interpretación del efecto parcial de *habit* en el *precio*:

$$\Delta \log\widehat{(precio)} = [-0.545 + 2(0.062)habit]\Delta habit$$

$$\%\Delta\widehat{precio} = 100[-0.545 + 2(0.062)habit]\Delta habit$$

  - Incremento de habit de 5 a 6 es 7.5%, de 6 a 7 es 19.9% y de 7 a 8 es 32.3%
  - Fuerte efecto creciente
  - Coeficiente de $\widehat{\beta}_2$ pequeño con grandes consecuencias ($2\widehat{\beta}_2$ no solo $\widehat{\beta}_2$)
  - Se puede comparar efecto parcial contra $\widehat{\beta}_1$ cuando $\beta_2 = 0$

Ejemplo. Elasticidad no constante

$\log(prc) = \beta_0 + \beta_1 \log(nox) + \beta_2[\log(nox)]^2 + \beta_3 crimen + \beta_4 habit + \beta_5 habit^2 + \beta_6 stratio + u$

- Si $\beta_2 = 0$, β_1 es elasticidad del *precio* con respecto a *nox*
- Si $\beta_2 \neq 0$, elasticidad del *precio* con respecto a *nox* depende de $\log(nox)$
- $\Delta \log(prc) \approx [\beta_1 + 2\beta_2 \log(nox)]\Delta \log(nox)$

$$\%\Delta prc \approx [\beta_1 + 2\beta_2 \log(nox)]\%\Delta nox$$

Se pueden tener otros términos polinomiales (aunque cuadrado es el más común)

- Ej. Función de costos total

$$costo = \beta_0 + \beta_1 cantidad + \beta_2 cantidad^2 + \beta_3 cantidad^3 + u$$

- No hay problema para estimarlo, se interpreta usando cálculo

6.2.3 Modelos con interacciones

Efecto parcial o (semi) elasticidad de la variable dependiente con respecto a una variable independiente puede depender de la magnitud de otra variable independiente

Ejemplo. Interacción en precios de casas

$$precio = \beta_0 + \beta_1 sqmts + \beta_2 habit + \beta_3 sqmts \cdot habit + u$$

- Efecto parcial de *habit* sobre *precio* (dejando otras variables fijas):

$$\frac{\Delta precio}{\Delta habit} = \beta_2 + \beta_3 sqmts$$

- Si $\beta_3 > 0$, una habitación extra incrementa más el precio de casas grandes

- Hay efecto interacción entre *sqmts* y *habit*

- Podemos probar diferentes hipótesis
  - $H_0 : \beta_3 = 0$
  - $H_0 : \beta_2 = 0, \beta_3 = 0$ (incluso si individuales no se rechazan)

- β_2 es efecto de *habit* en *precio* cuando $sqmts = 0$ (no es de interés)
  - Evaluar efecto parcial en valores relevantes de *sqmts*
    - Media, mediana, cuantiles bajo y alto
  - Reparametrizar

Podemos reparametrizar modelo para que coeficientes en tengan significado útil

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 \cdot x_2 + u$$

- Usamos medias poblacionales de x_1 y x_2: μ_1 y μ_2 (en la práctica, medias muestrales)

$$\begin{aligned}
y &= \alpha_0 + \delta_1 x_1 + \delta_2 x_2 + \beta_3 (x_1 - \mu_1) \cdot (x_2 - \mu_2) + u \\
&= \alpha_0 + \delta_1 x_1 + \delta_2 x_2 + \beta_3 (x_1 x_2 - x_1 \mu_2 - x_2 \mu_1 + \mu_1 \mu_2) + u \\
&= \underbrace{\alpha_0 + \beta_3 \mu_1 \mu_2}_{\beta_0} + \underbrace{(\delta_1 - \beta_3 \mu_2)}_{\beta_1} x_1 + \underbrace{(\delta_2 - \beta_3 \mu_1)}_{\beta_2} x_2 + \beta_3 x_1 x_2 + +u
\end{aligned}$$

- $\delta_2 = \beta_2 + \beta_3 \mu_1$: Efecto parcial de x_2 en y al valor medio de x_1
- Sustraer medias antes de definir la interacción
- Coeficientes de x_1 y x_2 tienen interpretación útil
  - Además, obtenemos errores estándar para el efecto parcial en el valor medio
- Podemos usar otros valores de interés (no solo μ_1 y μ_2)
  - Ej. Medianas

6.3 Bondad de ajuste y seleccion de regresores

Usar R^2 para escoger un conjunto de variables independientes no es recomendable

- Ej. Modelos sin sentido
- Ej. R^2 artificialmente alta

Un bajo poder explicativo ($R^2 \approx 0$) no tiene que ver con una estimación sesgada de β_j

- $R^2 \approx 0$ no implica que factores en u se correlacionan con variables independientes
- Falta de sesgo viene de RLM.4 no de R^2

Podemos tener estimados de efectos parciales precisos (sin n grande) aunque controlemos por pocos factores ($R^2 \approx 0$)

- Podemos contrarrestar una varianza del error grande con n grande

Lo que es útil es el <u>cambio relativo</u> en R^2 cuando agregamos variables en la ecuación

- Estadístico F usa R^2_r y R^2_{sr}

$\mathrm{R}^2 \approx 0$ implica que es difícil predecir y

6.3.1 R^2 ajustada (estadístico)

R^2 de la población: $\rho^2 = 1 - \frac{\sigma_u^2}{\sigma_y^2}$

- Hasta ahora $\mathrm{R}^2 = 1 - \frac{\mathrm{SCR}}{\mathrm{SCT}} = 1 - \frac{\mathrm{SCR}/n}{\mathrm{SCT}/n}$
- Sabemos que SCR/n y SCT/n son estimadores sesgados de σ_u^2 y σ_y^2
- En ambos casos, podemos corregir por los grados de libertad

R^2 ajustada o $\overline{\mathrm{R}}^2$:

$$\overline{\mathrm{R}}^2 = 1 - \frac{\mathrm{SCR}/(n-k-1)}{\mathrm{SCT}/(n-1)} = 1 - \frac{\widehat{\sigma}^2}{\mathrm{SCT}/(n-1)} = 1 - (1-\mathrm{R}^2)\left(\frac{n-1}{n-k-1}\right)$$

- Depende explícitamente de k
- No necesariamente es mejor estimador de ρ^2 que R^2
  - División de estimadores insesgados no es insesgada

Utilidad: Penaliza por agregar variables al modelo

- Si agregamos una variable al modelo,
  - SCR baja por lo que R^2 sube
  - SCR baja y $n-k-1$ baja por lo que $\mathrm{SCR}/(n-k-1)$ puede subir o bajar

$\overline{\mathrm{R}}^2$ da una respuesta distinta de pruebas t y F estándar porque $\overline{\mathrm{R}}^2$ sube si y solo si

- $\mathrm{t} > 1$ para nueva variable
- $\mathrm{F} > 1$ para conjunto de nuevas variables

Con n pequeño y k grande, $\overline{\mathrm{R}}^2 << \mathrm{R}^2$

- Ej. $\mathrm{R}^2 = 0.3$, $n = 51$, $k = 10 \implies \overline{\mathrm{R}}^2 = 1 - 0.7(\frac{50}{40}) = 0.125$

Si $\mathrm{R}^2 \approx 0$ y $n-k-1$ pequeño, $\overline{\mathrm{R}}^2 < 0$

- Ej. $\mathrm{R}^2 = 0.1$, $n = 51$, $k = 10 \implies \overline{\mathrm{R}}^2 = 1 - 0.9(\frac{50}{40}) = -0.125$
  - Muy mal ajuste dados los grados de libertad

A veces se reportan R^2 y $\overline{\mathrm{R}}^2$, y a veces solo se reporta $\overline{\mathrm{R}}^2$

- Para el estadístico F, se utiliza R^2 no $\overline{\mathrm{R}}^2$

6.3.2 R^2 permite escoger entre modelos no anidados

Estadístico F solo nos permite probar modelos anidados

- Modelo restringido es caso especial del modelo no restringido

En modelos no anidados, ninguno es un caso especial del otro

> Ejemplo. Como *hruns* y *carrbat* están altamente correlacionadas, podemos escoger entre
>
> $\log(salar) = \beta_0+\beta_1 ansliga+\beta_2 promjueg+\beta_3 prombat+\beta_4 hruns+u \to \overline{\mathrm{R}}^2 = 0.6211$
>
> $\log(salar) = \beta_0+\beta_1 ansliga+\beta_2 promjueg+\beta_3 prombat+\beta_5 carrbat+u \to \overline{\mathrm{R}}^2 = 0.6226$
>
> - ¿Cuál escogemos?

Comparar $\overline{\mathrm{R}}^2$ es útil para escoger entre diferentes conjuntos de variables independientes no anidadas cuando representan formas funcionales diferentes

> Ejemplo.
>
> $$\begin{aligned} investdes &= \beta_0 + \beta_1 \log(ventas) + u \\ investdes &= \beta_0 + \beta_1 ventas + \beta_2 ventas^2 + u \end{aligned}$$
>
> - Ambos modelos capturan rendimientos decrecientes
>   - Uno con log y el otro con término cuadrático
> - Primer modelo es más parsimonioso (menos parámetros)
>   - R^2: 0.061 vs 0.148
>   - Injusto comparar R^2's porque no penaliza modelos complicados
>   - $\overline{\mathrm{R}}^2$: 0.030 vs 0.090

Limitación:

- $\overline{\mathrm{R}}^2$ no sirve para escoger entre diferentes formas funcionales de variable dependiente
  - Porque explica variación de dos variables dependientes diferentes
  - Ej. y vs $\log(y)$, $\mathrm{Var}\left[\log(y)\right] < \mathrm{Var}\left(y\right)$ y $\overline{\mathrm{R}}^2_{\log(y)} > \overline{\mathrm{R}}^2_y$
- En ese caso, usar significancia estadística de variable o interpretación de coeficientes

6.3.3 Controlar por muchos factores

Para no omitir factores importantes en un modelo <u>sobre controlamos</u> por muchas variables

- Porque pudieran correlacionarse con las variables independientes
- Pero al dar demasiado peso a R^2 o $\overline{\mathrm{R}}^2$ podemos incluir factores que no deberíamos

Recordar interpretación ceteris paribus de RLM para evitar ese error

- No controlar por variable que necesitamos que cambie
  - Ej. Consumo de cerveza e impuesto alcohol
- No controlar por variable que es parte de y
  - Ej. Gasto en salud y visitas al doctor

La pregunta de investigación puede ayudar a decidir variables a incluir

- Diferentes modelos sirven diferentes objetivos
- Algunas variables independientes no se deberían incluir en el modelo
  - Aunque estén correlacionadas con y

6.3.4 Agregar regresores para reducir la varianza del error

Incluir una variable independiente puede tener efectos opuestos

- Exacerbar multicolinealidad
- Reducir varianza del error

En general, no sabemos qué efecto dominará, pero hay un caso que es claro:

- Siempre incluir variables independientes que
  - Afecten a y
  - No correlacionadas con la variable independiente de interés
- Razones:
  - No inducirán multicolinealidad
  - Reducirán la varianza del error
- Implicación: Estimadores más precisos para variable de interés
- Ej. Variable de interés se asigna aleatoriamente
  - $y = calif$, $x_1 = beca$, x's: $prepa$, $admision$, $familia$
  - Estimador ya sería insesgado
  - Pero queremos que tenga menor varianza muestral
- Desafortunadamente, dichas variables no son fáciles de obtener en ciencias sociales