WOMEN IN ECONOMICS - PERÚ
UNIVERSITARIOS SIN FRONTERAS

MICROECONOMETRÍA
Modelo de Regresión Simple
Sesión N°01

Docente: Carolina Saavedra

Asistenta de docencia: Emily Saavedra

Horario: Martes 8:00 p.m. – 10:00 p.m.

Recordando…

Ejemplo 1: Una ecuación simple sobre el salario

Un modelo en el que se relaciona el salario de una persona con la educación observada y con otros factores no observados es:

$$salario = \beta_0 + \beta_1 educación + \mu \quad \text{(I)}$$

Si $salario$ se mide en soles por hora y $educación$ se mide en años de educación, entonces β_1 mide la variación en el salario por hora por cada año más de educación, cuando todos los demás factores permanecen constantes (*ceteris paribus*). Entre estos factores se encuentran experiencia laboral, capacidades innatas, antigüedad en el empleo actual, ética laboral y otras influencias.

Notar que:

$$Y = \beta_0 + \beta_1 X + \mu \quad \text{(II)}$$

- La linealidad de esta ecuación implica que todo cambio de X en una unidad tiene siempre el mismo efecto sobre Y, sin importar el valor inicial de X. En muchas aplicaciones de la economía esto no es muy realista.
- El problema más difícil de abordar es si el modelo dado por la ecuación (II) en realidad permite formular conclusiones *ceteris paribus* acerca de cómo afecta X a Y.
- Problema de la causalidad: ¿Cómo esperar conocer el efecto ceteris paribus de x sobre y, cuando todos los demás factores permanecen constantes, si se ignoran esos otros factores?
- Veremos que… la única manera de obtener estimadores confiables de β_0 y β_1, a partir de los datos de una muestra aleatoria, es haciendo una suposición que restrinja la manera en que la variable no observable μ está relacionada con la variable explicativa X. Sin esta suposición, no es posible estimar el efecto *ceteris paribus.*

Antes de establecer esta suposición clave acerca de la relación entre X y μ se puede hacer una suposición acerca de μ.

$$E(\mu) = 0 \quad \text{(III)}$$

Ahora, volvamos al supuesto crucial sobre la manera en que están relacionadas μ y X. Una
medida natural de la relación entre dos variables aleatorias es el coeficiente de correlación. Si μ y X no están correlacionadas, entonces, como variables aleatorias, no están relacionadas linealmente. Suponer que μ y X no están correlacionadas es un avance para definir el sentido en el que μ y X estarán relacionadas en la ecuación (II). Sin embargo, el avance no es suficiente.

El supuesto crucial es que el valor promedio de μ no depende del valor de X. Este supuesto se expresa como:

$$E(\mu|X) = E(\mu) \quad \text{(IV)}$$

La ecuación (IV) indica que el valor promedio de los factores no observables es el mismo en todas las fracciones de la población determinados por los valores de X y que este promedio común es necesariamente igual al promedio de μ en toda la población. Cuando se satisface el supuesto (ecucación IV) se dice que u es media independiente de X.

Combinando la independencia de la media con el supuesto (III), se obtiene el supuesto de media condicional cero, es decir:

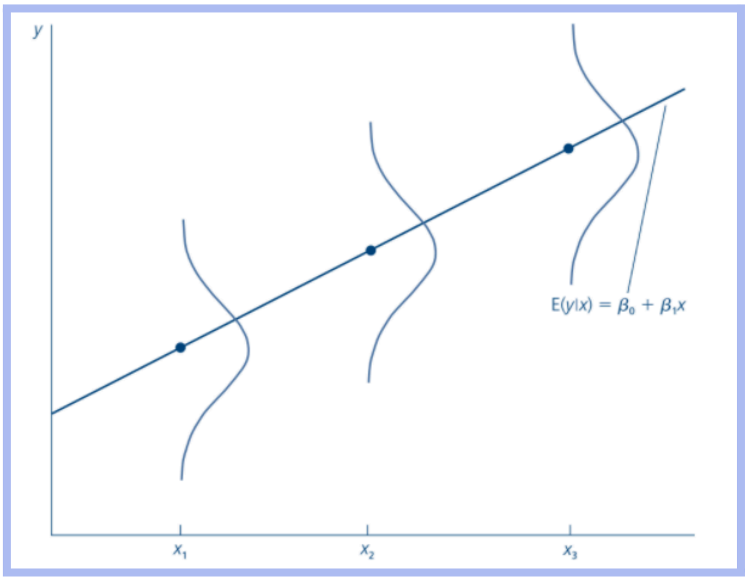
$$E(\mu|X) = 0 \quad \text{(V)}$$

El supuesto de media condicional cero proporciona otra interpretación de β_1 que suele ser útil. Tomando el valor esperado de (II) condicionado a X, usando $E(\mu|X) = 0$, se tiene:

$$E(Y|X) = \beta_0 + \beta_1 X \quad \text{(VI)}$$

La ecuación (VI) muestra que la función de regresión poblacional (FRP), $E(Y|X)$, es una
función lineal de X. La linealidad significa que por cada aumento de una unidad en X el valor esperado de Y se modifica en la cantidad β_1. Dado cualquier valor de X, la distribución de Y está centrada en $E(Y|X)$, como se ilustra en la siguiente figura:

Figura 01: $E(Y|X)$ como función lineal de X



Fuente: Wooldridge (2010)

Es importante entender que la ecuación (VI) dice cómo varía el valor promedio de Y de acuerdo con la variación de X; esta ecuación no dice que Y sea igual a $\beta_0 + \beta_1 X$ para cada una de las unidades de la población. La FRP da una relación entre el promedio de Y y diferentes valores de X.

Dado el supuesto de media condicional cero $E(\mu|X) = 0$, es útil ver la ecuación (II) como
una que divide a Y en dos componentes. A la parte $\beta_0 + \beta_1 X$, que representa $E(Y|X)$, se le llama parte sistemática de Y, es decir, es la parte de Y explicada por X y a μ se le llama la parte no sistemática, o la parte de Y que no es explicada por X.

Ahora bien, usaremos los supuestos (III) y (IV) para obtener estimadores de β_0 y β_1 a partir de una muestra aleatoria de datos dada.

Obtención de las estimaciones de mínimos cuadrados ordinarios

1	Estimación por Mínimos Cuadrados Ordinarios (MCO): Intercepto $SSE = \sum_{i=1}^{n}[Y_i - \hat{a} - \hat{b}X_i]^2$ $\frac{\partial SSE}{\partial \hat{a}} = -2\sum_{i=1}^{n}[Y_i - \hat{a} - \hat{b}X_i] = 0$ $\rightarrow -\sum Y_i - n\hat{a} + \hat{b}\sum X_i = 0 \rightarrow \hat{a}n = \sum Y_i - \hat{b}\sum X$ $\Rightarrow \hat{a} = \overline{Y} - \hat{b}\overline{X}$

2	**Estimación por Mínimos Cuadrados Ordinarios (MCO): Pendiente** $SSE = \sum_{i=1}^{n} [Y_i - \hat{a} - \hat{b}X_i]^2$ $\frac{\partial SSE}{\partial \hat{b}} = -2\sum_{i=1}^{n} [Y_i - \hat{a} - \hat{b}X_i] \times X_i = 0$ ... y al despejar $\hat{b}$ resulta en : $\hat{b} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(X_i - \overline{X})^2}$
3	Para asegurarnos de tener un mínimo local ● La condición suficiente de segundo orden para tener un mínimo local es que la matriz Hessiana debe ser positiva-definida, es decir, esta matriz debe cumplir: i) determinante sea positivo tanto para la matriz hessiana como sus submatrices, y ii) debe ser simétrica. ● La matriz Hessiana para el modelo MCO, es la matriz de segundas derivadas: $H = \begin{bmatrix} 2n & 2\sum_{i=1}^{n} X_i \\ 2\sum_{i=1}^{n} X_i & 2\sum_{i=1}^{n} X_i^2 \end{bmatrix}$ Determinante H_1: $2n > 0$ Determinante H: $4n\sum X_i^2 - 4[\sum X_i]^2 > 0$
4	**Derivación** $\frac{\partial SCR}{\partial \hat{a} \partial \hat{a}} = -2\left(\sum -1\right) = 2n$ $\frac{\partial SCR}{\partial \hat{b} \partial \hat{a}} = (-1)(-2)\left(\sum X_i\right) = 2\sum X_i$ $\frac{\partial SCR}{\partial \hat{a} \partial \hat{b}} = (-1)(-2)\left(\sum X_i\right) = 2\sum X_i$ $\frac{\partial SCR}{\partial \hat{b} \partial \hat{b}} = (-1)(-2)\left(\sum X_i^2\right) = 2\sum X_i^2$
5	**La bondad de ajuste del modelo (i)** Mientras mayor es la proporción de la varianza de Y explicada por el modelo, mayor será la bondad de ajuste del modelo (R^2) $SCT \equiv \sum [Y_i - \overline{Y}]^2$ $SCE \equiv \sum [\hat{Y}_i - \overline{Y}]^2$ $SCR \equiv \sum [Y_i - \hat{Y}]^2$ SCT: es la suma de cuadrados totales y representa la variación total de la dependiente en la muestra que se esta analizando SCE: es la suma de cuadrados explicada o en otras palabras la variación de la variable predicha con el modelo estimado. SCR: es la suma de cuadrados de la diferencia entre el valor observado y el valor que predice mi modelo (errores).

6	**La bondad de ajuste del modelo (ii)** Así, nosotros podemos decir los siguiente: SCT = SCE + SCR Divido todo entre la SCT y obtenemos $\frac{SCT}{SCT} = \frac{SCE}{SCT} + \frac{SCR}{SCT}$ $1 = R^2 + \frac{SCR}{SCT}$ entonces $R^2 = 1 - \frac{SCR}{SCT}$ En otras palabras, el coeficiente de determinación o R^2 es el ratio entre la variación explicada y la variación total, es decir, que proporción de la variación de la dependiente es explicada por la variable explicativa
7	**La bondad de ajuste del modelo (iii)** Un aspecto a tomar en cuenta es que el coeficiente de determinación si bien indica que proporción de la variabilidad de la dependiente estamos explicando, cuando se realiza análisis con datos de corte transversal este no es muy alto, razón por la cual no se le debe dar tanto peso al tamaño del R^2 cuando se realiza análisis con este tipo de datos. Por el contrario, cuando se trabaja con bases de datos temporales, si se debe dar énfasis al tamaño del R^2 Por otro lado el R^2 también es igual al cuadrado del coeficiente de correlación de Pearson entre las variables dependiente e independiente. A medida que el valor del R^2 es mayor, el ajuste de la recta de regresión a los datos observados es mejor, puesto que estamos explicando mejor la variabilidad.
8	**Demostración** SCT = SCE + SCR $\sum(Y_i - \bar{Y})^2 = \sum(Y_i - \hat{Y} + \hat{Y} - \bar{Y})^2 = \sum(\mu_i + (\hat{Y} - \bar{Y}))^2$ $= \sum(\mu_i^2 + (\hat{Y} - \bar{Y})^2 + 2\mu_i(\hat{Y} - \bar{Y})) = \sum \mu_i^2 + \sum(\hat{Y} - \bar{Y})^2 + 2\sum \mu_i(\hat{Y} - \bar{Y})$ $SCT = \sum \mu_i^2 + \sum(\hat{Y} - \bar{Y})^2 = SCR + SCE$

Propiedades del estimador MCO

En general, se utiliza el estimador que posee mejores propiedades que los restantes, como insesgadez, eficiencia y consistencia. El valor de un estimador proporciona lo que se denomina en estadística una estimación puntual del valor del parámetro en estudio.

Insesgadez

$$E(\hat{\vartheta}) = \theta$$

En caso contrario se dice que el estimador es sesgado. Se llama sesgo a $B(\hat{\vartheta}) = \theta.\ E(\hat{\vartheta})$

Un estimador insesgado es aquel cuya esperanza matemática coincide con el valor del parámetro que se desea estimar. En caso de no coincidir se dice que el estimador tiene sesgo. Matemáticamente se expresa de la siguiente forma:

Consistencia

Un estimador es consistente si converge en probabilidad al parámetro a estimar . Esto es:

$$(\left|\hat{\vartheta} - \theta\right| < \varepsilon) \ = 1$$

Eficiencia

Un estimador es eficiente u óptimo cuando posee varianza mínima o bien en términos relativos cuando presenta menor varianza que otro . Puede plantearse también en términos más coherentes de Error Cuadrático Medio (ECM). Tendríamos que :

$$ECM(\hat{\vartheta}) = E[\left(\hat{\vartheta} - \theta\right)^2 = D^2\left[\hat{\vartheta}\right] + B^2\left[\hat{\vartheta}\right]$$

Resumen de la sesión

Supuestos fundamentales de un modelo de regresión simple

Sea la siguiente ecuación $Y_i = \beta_0 + \beta_1 X_i + \ \mu_i$, se debe cumplir:

- Linealidad en los parámetros.
- Los datos provienen de una muestra aleatoria.
- Variación en las variables explicativas X_i
- El valor esperado de los errores sea 0 para cualquier valor de X_i, es decir, $E(\mu_i | X_i) \ = 0$
- Los errores de la regresión son Homocedasticos, es decir que tienen varianza constante (σ^2) y media cero $(\mu \ = 0)$
- No existe autocorrelación serial, es decir, no hay covarianza o correlación entre los
- errores de diferentes periodos: $cov\left(\mu_i, \mu_j\right) = 0$

Interpretación de los parámetros estimados

Especificación	Expresión	Interpretación de $\hat{\beta}_2$
Nivel – Nivel	$y_i = \beta_1 + \beta_2 X_{2i} + \mu_i$	Incremento de unidades en “y” cuando aumenta 1 unidad la “x” (ambas en sus unidades de medida originales)

Nivel – Log	$log(y_i) = \beta_1 + \beta_2 X_{2i} + \mu_i$	$\hat{\beta}_2 * 100$ = Incremento porcentual de "y" cuando aumenta una unidad la "x"
Log – Nivel	$y_i = \beta_1 + \beta_2 log(X_{2i}) + \mu_i$	$\frac{\hat{\beta}_2}{100}$ = Incremento en unidades de "y" cuando aumenta una unidad la "x"
Log – Log	$log(y_i) = \beta_1 + \beta_2 log(X_{2i}) + \mu_i$	Incremento porcentual de "y" cuando aumenta un "1%" la "x"

Bibliografía recomendada

1. Introductory Econometrics. A modern Approach- Wooldridge, 4° edition
2. Mastering Metrics: The Path from Cause to Effect – Angrist y Pischke (Próximas clases)
3. UCLA: Statistical Methods and Data Analytics. Ver en: https://stats.oarc.ucla.edu/other/annotatedoutput/