Econometría I - 2023
Facultad de Ciencias Económicas
Universidad Nacional de La Plata

Trabajo Práctico 1

Modelo Lineal con dos variables

Contenidos: Modelo Lineal con 2 variables, interpretación de los parámetros poblacionales lineales, log-log, log-lin, lin-lin, unidades de medida, estimadores Mínimos Cuadrados Ordinarios (MCO), propiedades algebraicas de los estimadores MCO, supuestos clásicos, propiedades estadísticas de los estimadores MCO, test "t", p-valores, coeficiente de determinación o R-cuadrado.

Fecha de entrega: jueves 7 de septiembre. Ver reglas de formato y presentación en https://econometria1unlp.weebly.com/praacutectica.html.

Parte I

1. Teoría de la producción

Existen distintos modelos o formas de representar las funciones de producción en la microeconomía. Cada una de ellas tiene alguna característica que la distingue de las otras en cuanto a su interpretación analítica y económica. Vamos a trabajar con la versión econométrica de estos modelos, donde explícitamente damos cuenta de que la función de producción en cuestión representa a una muestra particular y que existen factores inobservables. Sean Y_i las unidades producidas por semana de un bien por la firma i, L_i las horas de trabajo semanales afectadas a la producción del bien Y_i y μ_i factores inobservables que determinan la producción más allá de las horas de trabajo. Considerar las siguientes formas funcionales o especificaciones alternativas:

(1) $Y_i = \alpha_1 + \beta_1 L_i + \mu_i$ (2) $\ln Y_i = \alpha_2 + \beta_2 \ln L_i + \mu_i$

En base a estas ecuaciones se pide lo siguiente:

a) Indicar de qué tipo de modelo se trata de acuerdo a cómo están expresadas las variables (lineales, logarítmicas).

b) Para cada modelo, interpretar económicamente los parámetros denotados con la letra β, teniendo en cuenta las unidades de medida y la forma funcional en que están expresadas las variables.

c) Considerar el parámetro α. ¿Tiene siempre una interpretación económica? Justificar.

d) Para el modelo (2), explicar qué podríamos aprender acerca de la tecnología de producción si contáramos con una estimación de los valores de β.

2. El ahorro como determinante de la inversión

En este ejercicio vamos a estudiar la relación entre la tasa de inversión y la tasa de ahorro replicando un clásico en la literatura como es el trabajo de Martin Feldstein y Charles Horioka (1980). Estos autores estudian el financiamiento externo de la economía a través del mercado de capitales. Según la teoría económica, la inversión es financiada por el ahorro interno y externo: si la inversión es independiente del nivel de ahorro interno entonces se considera que hay perfecta movilidad de capitales desde el exterior; por otro lado, si un peso adicional de inversión es totalmente financiado por un peso adicional de ahorro interno, entonces se considera que la economía está cerrada al mercado de capitales extranjeros. Una versión simplificada del modelo propuesto por los autores es la siguiente:

(1) $$I_i = \alpha + \beta S_i + \mu_i$$

donde I_i (variable explicada) es la inversión medida en millones de dólares y S_i (variable explicativa) es el ahorro interno medido en millones de dólares. Suponer los siguientes valores de los parámetros: $\alpha = 0.015$ y $\beta = 0.957$

a) Indicar de qué tipo de modelo se trata de acuerdo con cómo están expresadas las variables.

b) Interpretar analítica y económicamente el parámetro α.

c) Calcular el efecto marginal del ahorro sobre la inversión e interpretar su significado económico.

3. Federalismo Fiscal

Suponer que fueron contratados por el Ministro de Economía con el objetivo de estudiar cómo es la relación entre el gasto público *per cápita* y la densidad poblacional a nivel de cada municipio de la provincia, de manera de poder mejorar la asignación del presupuesto provincial. En la literatura de las Finanzas Públicas existe un gran interés en el coeficiente de regresión entre el gasto público per cápita y el número de habitantes, ya que ese coeficiente informa sobre la existencia de economías o deseconomías de escala (y costos de congestión) en el gasto público. El modelo propuesto para su estudio es el siguiente:

(1) $$ln(Gpubpc_i) = \alpha + \beta Dens_i + \mu_i$$

donde $ln(Gpubpc_i)$ es el logaritmo del gasto público *per cápita* del municipio i, $Dens_i$ es la densidad poblacional medida como número de habitantes por km^2 y μ_i es un término aleatorio de ese mismo municipio.

a) Indicar de qué tipo de modelo se trata de acuerdo con cómo están expresadas las variables.

b) Suponiendo que $\beta = -0.0013$, obtener analíticamente el efecto marginal de un cambio en la densidad del municipio sobre el gasto público *per cápita*. Interpretar económicamente.

c) El municipio de La Plata tiene una extensión de 926 Km^2 y se estima que entre 2019 y 2025 su población pasará de 708.733 a 738.505 habitantes. Suponiendo nuevamente que $\beta = -0.0013$ ¿Cuál es el cambio en el gasto público per cápita que deberíamos esperar como consecuencia del crecimiento poblacional en este municipio?

d) ¿Como se interpretaría económicamente β si la variable $Dens_i$ estuviese también medida en logaritmos? ¿De qué tipo de modelo se trataría en ese caso?

Parte II [1][2]

4. Modelo lineal con dos variables

Con los siguientes datos de ahorro S_i e ingreso I_i (medidos en pesos) para 15 familias:

Hogar (*i*)	Ahorro (*S*)	Ingreso (*I*)
1	97	141
2	46	145
3	125	209
4	81	132
5	68	137
6	147	205
7	12	61
8	84	136
9	105	175
10	93	140
11	15	41
12	112	158
13	14	42
14	61	78
15	59	130

a) Estimar la covarianza y la correlación entre el ahorro y el ingreso de las familias. Interpretar.

b) Suponer que la relación entre el ahorro y el ingreso de las familias puede representarse mediante el siguiente modelo

$$S_i = \alpha + \beta I_i + \mu_i \qquad i = 1, \ldots, n$$

en donde μ_i es un término aleatorio no observable. Obtener y estimar los parámetros α y β de este modelo mediante el método de Mínimos Cuadrados Ordinarios (MCO). Proporcionar una interpretación económica para las estimaciones obtenidas.

c) *Propiedades algebraicas de los estimadores MCO*. Utilizar las condiciones de primer orden del método de MCO para demostrar analíticamente que la suma de los errores de todas las familias es cero. Adicionalmente, verificar la demostración numéricamente usando los datos provistos en el enunciado.

d) Construir un gráfico de dispersión de los datos (nube de puntos) ubicando en el eje de ordenadas los valores del ahorro y en el de abscisas el ingreso. En el mismo gráfico, dibujar la recta de regresión estimada en el inciso b. [Nota: no es un dibujo a mano alzada, sino uno que se corresponda con los datos proporcionados].

---

[1] Para todos los ejercicios de la parte II de este Trabajo Práctico se puede utilizar calculadora o planilla de cálculo, pero <u>no STATA.</u>

[2] Adjuntar en un archivo separado la planilla de Excel o la hoja de cálculos utilizada (en forma prolija).

5. Cambios en las unidades de medida en que se expresan las variables

a) Estimar el coeficiente β del modelo de ahorro e ingreso del ejercicio anterior, pero con el ahorro medido en centavos. Interpretar el nuevo valor estimado y compararlo con el obtenido antes del cambio en la unidad de medida.

b) Mostrar formalmente que el coeficiente de correlación muestral no cambia cuando la variable S (dependiente) se multiplica por una constante $k > 0$. Mostrar formalmente cómo cambia el estimador MCO de β ante esa misma transformación.

c) Repetir (a) para el mismo cambio pero aplicado a los ingresos (variable explicativa).

6. Test de hipótesis sobre los parámetros del modelo lineal

En todos los casos realizar un test de hipótesis y explicar detalladamente paso por paso: cuáles son las hipótesis nula y alternativa, cuál es el estadístico de prueba y su distribución bajo la hipótesis nula, la regla de decisión y el resultado del test. Utilizar un nivel de confianza del 95%.

El ahorro como determinante de la inversión: retomar las interpretaciones económicas del modelo sobre ahorro e inversión agregados visto en el ejercicio 2, suponiendo que se cuenta con las siguientes estimaciones por MCO para el parámetro β basadas en una muestra de 100 países: $\hat{\beta}$ = 0.957 y su error estándar de $SE(\hat{\beta})$ = 0.015, donde tanto ahorro e inversión están medidos en millones de dólares. Con esta información se pide:

a) Evaluar la hipótesis nula de que el ahorro no es relevante para explicar la inversión. Utilizar una hipótesis alternativa a dos colas.

b) Evaluar la hipótesis nula de que $\beta = 1$ contra la hipótesis alternativa bilateral. Interpretar económicamente el resultado del test.

Federalismo Fiscal: Suponer que se estimó el modelo planteado en el ejercicio 3 con datos de 134 municipios de la provincia de Buenos Aires. Utilizando el método de MCO se obtuvo $\hat{\beta}$ = -0.0013 y su error estándar $SE(\hat{\beta})$ = 0.0006. En base a esta información resolver los siguientes incisos planteándolos como pruebas de hipótesis:

c) ¿Es la densidad poblacional una variable relevante para explicar el gasto municipal per cápita? Realizar la prueba contra una hipótesis alternativa bilateral.

d) Evaluar la misma hipótesis nula del inciso anterior contra una hipótesis alternativa que plantea la existencia de economías de escala en el gasto público ¿Qué ocurre si en cambio se utiliza un nivel de confianza del 99%?

7. Propiedades estadísticas de los estimadores MC

La teoría nos da una base sólida para entender las propiedades estadísticas de los estimadores mínimo cuadráticos y cómo esas propiedades dependen de los distintos supuestos. En este ejercicio vamos a hacer una simulación en STATA. Vamos a proceder como si conociéramos el verdadero modelo, y generaremos nuestros propios datos a partir de ese modelo para observar cómo los estimadores mínimo cuadráticos se desempeñan en distintos escenarios.

Supongamos que conocemos el modelo poblacional verdadero dado por:

$$Y_i = \alpha + \beta X_i + \mu_i$$

Con parámetros α=15 y β=2.

Supongamos que tenemos una muestra de 25 observaciones y que los regresores fijos toman los valores de los números enteros entre 1 y 25. Es decir que $X_1 = 1,\ X_2 = 2, \ldots ,\ X_{25} = 25$.
Sabemos que μ_i satisface los supuestos clásicos y, adicionalmente, se cumple que tiene distribución normal de tal forma que $\mu_i \sim N(0,1)$ con i=1, ..., 25.

a) **Insesgadez del estimador de MCO**
Generar la muestra aleatoria 10 veces y estimar los parámetros poblaciones en cada una de dichas muestras por MCO. Luego, hacer un promedio simple entre las 10 estimaciones de α y las 10 estimaciones de β. Relacionar los resultados obtenidos con la propiedad de insesgadez de los estimadores de los parámetros.

Observación: no se debe fijar una semilla (seed) al generar cada muestra aleatoria, ya que en dicho caso se estaría generando siempre la misma muestra.

b) **Sesgadez del estimador de β al omitir la constante**
Volver a generar la muestra aleatoria 10 veces usando el mismo modelo poblacional que en el inciso (a) pero estimar el parámetro β utilizando un modelo sin constante en cada una de dichas muestras. Luego, hacer un promedio entre las 10 estimaciones de β. ¿Se mantiene la insesgadez de $\hat{\beta}$? ¿Por qué?

Graficar sólo una de las muestras generadas en el plano X-Y mediante un gráfico de dispersión. Trazar la recta de regresión estimada con el modelo con ordenada y con el modelo sin ordenada.

c) **La heterocedasticidad no está relacionada con la insesgadez**
Supongamos ahora que $\mu_i \sim N(0, i^2)$. Es decir, $\sigma_{\mu_1} = 1,\ \sigma_{\mu_2} = 2, \ldots ,\ \sigma_{\mu_{25}} = 25$.
Repetir la consigna del inciso (a). ¿Se mantiene la insesgadez de los estimadores? ¿Por qué?

Graficar sólo una de las muestras generadas en el plano X-Y. ¿Qué tiene de particular el gráfico? ¿Por qué?

d) **No es necesario suponer que $\mathbf{E(\mu_i) = 0}$ para tener insesgadez del estimador de la pendiente, solo que $\mathbf{E(\mu_i)}$ es constante.**
Supongamos ahora que $\mu_i \sim N(4,1)$. Es decir, volvemos al supuesto de homocedasticidad pero ahora suponemos que la media del error es igual a 4.

Repetir la consigna del inciso (a). ¿Se mantiene la insesgadez del estimador de β? ¿Ocurre lo mismo con el estimador de α? ¿Por qué?

e) **La insesgadez del estimador de MCO no depende del supuesto de normalidad de los residuos**.
Supongamos que los residuos tienen distribución uniforme de manera que $\mu_i \sim U[-5,5]$. Observar que se cumple que $E(\mu_i) = 0$.

Repetir la consigna del inciso (a). ¿Se mantiene la insesgadez de los estimadores de α y de β? ¿Por qué?

f**) Bajo el supuesto de normalidad de los residuos, los estimadores de los parámetros también tienen distribución normal.**
Volvamos al supuesto inicial de que $\mu_i \sim N(0,1)$. Generar la muestra aleatoria 1000 veces y guardar los coeficientes estimados $\hat{\beta}$. Estimar la función de densidad de dichos coeficientes a través de un histograma. ¿Qué forma tiene dicho gráfico y por qué?

Parte III[3]

8. Problema empírico: estimación de una curva de Engel

El objetivo de este ejercicio consiste en introducir el análisis de regresión bivariado a nivel empírico, es decir, a ejemplos concretos usando datos reales, y aprender los comandos básicos del paquete econométrico Stata.

Se estudiará la relación entre el gasto en alimentos, que es una aproximación del consumo y el ingreso de las familias. Las relaciones entre el consumo de un determinado bien y el ingreso se conocen con el nombre de "curvas de Engel".

Este ejercicio consiste en la estimación de la curva de Engel del consumo de alimentos de Argentina en su versión más simple. La base de datos que se empleará (*gasto-ingreso.dta*) fue obtenida a partir de los resultados de la Encuesta Nacional de Gasto de los Hogares (ENGH) realizada entre 2004 y 2005 por el INDEC. La base contiene información sobre el ingreso per cápita familiar (*ipcf*) y el gasto en alimentos per cápita (*gasto_alimentos*) de 2,659 hogares con ingresos y gastos en alimentos. Las dos variables se encuentran medidas en pesos corrientes.

El modelo propuesto es:

$$G_i = \alpha + \beta Y_i + u_i \qquad i = 1,..., 2,659$$

donde G, Y y u representan el gasto en alimentos per cápita familiar, el ingreso per cápita familiar y un término aleatorio no observable que suponemos cumple con los supuestos clásicos.

Se pide:

a) Realizar un análisis estadístico descriptivo básico de cada variable (computar y comentar los resultados para la media, la mediana, el rango, el coeficiente de variación y el desvío estándar).

b) Construir un diagrama de dispersión graficando el ingreso per cápita familiar en el eje de abscisas y el gasto en alimentos per cápita del hogar en el eje de ordenadas. Discutir.

c) Utilizando el método de Mínimos Cuadrados Ordinarios, estimar los parámetros del modelo propuesto más arriba, donde el gasto en alimentos per cápita es función únicamente del ingreso per cápita familiar. Añadir la recta resultante de esta estimación al gráfico del inciso b.

d) Interpretar económicamente el valor de los coeficientes obtenidos (considerar especialmente las unidades de medida).

e) Interpretación estadística de los coeficientes estimados. Evaluar la hipótesis nula de que el ingreso no es una variable relevante para explicar la demanda de alimentos de los hogares contra una hipótesis alternativa bilateral.

f) ¿Cuánto de la variabilidad del gasto per cápita en alimentos es explicada por la variabilidad del ingreso per cápita familiar?

[3] A diferencia de los ejercicios anteriores, los correspondientes a la parte III deben realizarse usando STATA