#NotesOnEconometrics

Economía UP

Índice

Capítulo 1

Bienvenido

#NotesonEconometrics es una iniciativa creada por alumnos de la Licenciatura en Economía de la Universidad Panamericana, con el objetivo de ofrecer recursos académicos accesibles y de alta calidad sobre econometría. Este proyecto forma parte de Economía UP, un esfuerzo estudiantil dedicado a fomentar el aprendizaje y el análisis económico de manera rigurosa y práctica, poniendo a disposición de los estudiantes herramientas que faciliten el estudio y la aplicación de conceptos econométricos fundamentales y avanzados.

En este sitio encontrarás contenido sintetizado de algunos de los libros más importantes en el campo de la econometría, como *Introductory Econometrics* de Jeffrey Wooldridge, *Basic Econometrics* de Damodar Gujarati e *Introduction to Econometrics* de G.S. Maddala. Estas obras han sido seleccionadas por su relevancia académica y su capacidad para explicar conceptos complejos de manera accesible. Además, el proyecto incorpora notas y recursos de cursos impartidos por profesores de la Universidad Panamericana, como Javier Alcántar Toledo y Eugenio Gómez Alatorre, cuyas enseñanzas han sido fundamentales para la formación en econometría de muchos estudiantes.

El contenido de **#NotesOnEconometrics** está diseñado para ayudar a los estudiantes a comprender y aplicar conceptos clave de econometría. A través de explicaciones claras, ejemplos prácticos y tutoriales de R y Python, buscamos no solo simplificar el aprendizaje de la teoría econométrica, sino también facilitar su aplicación en problemas del mundo real. Esto incluye desde los fundamentos de la regresión lineal hasta temas más avanzados, proporcionando herramientas que permitan a los usuarios desarrollar habilidades analíticas rigurosas y aplicarlas en investigaciones académicas o en el análisis de datos económicos.

Este proyecto está dirigido a estudiantes de economía, investigadores y cualquier persona interesada en aprender econometría de manera autodidacta o complementar sus estudios formales. Si estás cursando materias de econometría en la universidad o deseas fortalecer tus habilidades analíticas para enfrentar retos en el ámbito profesional, este sitio ha sido pensado para ti. La misión de **#NotesOnEconometrics** es hacer de la econometría un tema accesible y comprensible, apoyando a los estudiantes en su desarrollo académico y profesional a través de recursos cuidadosamente seleccionados y organizados.

Capítulo 2

Análisis de regresión

Econometría: Medición económica.

2.1 Metodología clásica

1. Planteamiento de teoría (hipótesis)
2. Especificación del modelo matemático
3. Especificación del modelo econométrico
4. Obtención de datos
5. Estimación de parámetros del modelo
6. Pruebas de hipótesis
7. Pronóstico (predicción)
8. Modelo para fines de control/política

Ej. Función consumo keynesiana:

$$c = \alpha + \beta y \quad \forall\ 0 < \beta < 1$$

Las relaciones entre variables económicas son **inexactas**, dada la injerencia de otras variables:

$$\underset{\text{Modelo econométrico}}{c = \alpha + \beta y + u}$$

$$\underset{\text{Variable aleatoria con propiedades probabilísticas}}{u = \text{Error (perturbación estocástica)}}$$

u incluye todos los factores que afectan **consumo** pero no están en la ecuación: tamaño de familia, edades, etc.

El modelo requiere ser estimado: obtener valores α y β a partir de **datos.**

Ej. Gasto en consumo personal

- Regresión: técnica estadística para el estudio de una **variable dependiente** que está en función de una o más **variables independientes.**
- Usando los datos de consumo y PIB de BUA se obtiene:

$$\hat{c} = -231.8 + 0.7194y$$

Donde:

- $\alpha = -231.8$
- $\beta = 0.7194$
- $\hat{c} =$ Consumo (estimado)
- $y =$ PIB.

La interpretación consiste en que un incremento de tasa en el ingreso incrementa (en promedio) el consumo en 0.72 USD.

Se debe probar si los valores estimados:

1. Son estadísticamente significativos ($\alpha, \beta \neq 0$)

2. Confirman la teoría (hipótesis) que está siendo probada ($0 < \beta < 1$)

Si el modelo confirma la teoría (hipótesis), se pueden **pronosticar** valores futuros de la variable dependiente.

Ej. Suponer un PIB esperado para 1995 de 6,000 mmd, ¿cuál es el pronóstico de consumo?

$$\hat{c} = -231.8 + 0.7194(6,000) = 4,084.6$$

Suponer que el gobierno considera que un gasto de \$4,000 mmd mantendrá la tasa de desempleo en 6.5%. ¿Cuál nivel de ingreso garantizará esta meta de consumo?

$$\begin{aligned}\hat{c} &= -231.8 + 0.7194y \\ 4,000 &= -231.8 + 0.7194y \\ 0.7194y &= 4,000 + 231.8 \\ y &= \frac{4,231.8}{0.7194} \\ y^* &= 5,882.40\end{aligned}$$

Un modelo estimado puede ser usado para fines de control o de política económica (fiscal y monetaria).

2.2 Probabilidad condicional

Suponer un país con una población de 60 familias. Se estudia el gasto en consumo familiar semanal (y) y el ingreso familiar semanal (x).

Se presenta la distribución del gasto en consumo (y) correspondiente a un ingreso fijo (x): la distribución condicional de y dada x.

Para $P(y|x = 80)$:

$$\begin{aligned}P(y = 55|x = 80) &= \frac{1}{5} \\ P(y = 60|x = 80) &= \frac{1}{5} \\ P(y = 65|x = 80) &= \frac{1}{5} \\ P(y = 70|x = 80) &= \frac{1}{5} \\ P(y = 75|x = 80) &= \frac{1}{5}\end{aligned}$$

Para cada distribución de probabilidad condicional de y, calculamos su **media (media condicional):**

$$E(y|x = 80) = 55\left(\frac{1}{5}\right) + 60\left(\frac{1}{5}\right) + 65\left(\frac{1}{5}\right) + 70\left(\frac{1}{5}\right) + 75\left(\frac{1}{5}\right) = 65$$

$$\mu_{y|x=100} = \frac{\sum_y y_i}{n} = \frac{462}{6} = 77$$

2.3 Función de regresión poblacional

Lugar geométrico de las medias condicionales de la variable dependiente para valores fijos de la variable independiente.

Se puede deducir que: $E(y|x_i) = f(x_i)$.

$$E(y|x_i) \rightarrow \text{Función de Regresión Poblacional (FRP)}$$

Forma funcional de la FRP:

$$E(y|x_i) = \alpha + \beta x_i \rightarrow \underset{y=ax+b}{\text{Ecuación de recta}}$$

$$\begin{array}{ccc} \alpha, \beta & \rightarrow & \text{Coeficientes de regresión} \\ \alpha & \rightarrow & \text{Intercepto} \\ \beta & \rightarrow & \text{Coeficiente de la pendiente} \end{array}$$

Objetivo: Estimar α y β con base en observaciones de x y y.
Esta desviación de un y_i alrededor de su valor esperado se expresa como:

$$u_i = y_i - E(y|x_i)$$
$$y_i = E(y|x_i) + u_i$$

Donde u_i es el término de error estocástico, que representa todas las variables omitidas que puedan afectar y, pero no están incluidas en el modelo de regresión.

$$y_i = \alpha + \beta x_i + u_i \quad \therefore \quad E(y|x_i) = \alpha + \beta x_i$$
$$\therefore \quad y_i = \alpha + \beta x_i + u_i \rightarrow \text{Ecuación de regresión}$$

Adicionalmente, si se toma $E(\cdot)$ a y_i, tenemos que:

$$\begin{aligned} y_i &= E(y|x_i) + u_i \\ E(y|x_i) &= E[E(y|x_i)] + E(u_i|x_i) \\ &= E(y|x_i) + E(u_i|x_i) \quad \therefore \\ E(u_i|x_i) &= E(y|x_i) - E(y|x_i) \quad \therefore \\ E(u_i|x_i) &= 0 \end{aligned}$$

2.4 Función de regresión muestral

A diferencia del caso anterior, en realidad se trabaja con muestras. Se estima la FRP con base en información **muestral.** Suponer que se extrae una muestra d ela población de 60 familias donde se estudia el gasto en consumo y el ingreso familiar. Muestra aleatoria de 10 observaciones.

[IMAGEN]

La función de regresión muestral (FRM) puede escribirse como:

$$\hat{y}_i = \hat{\alpha} + \hat{\beta} x_i$$

Donde:

- $\hat{y}_i \equiv y$ estimada
- $\hat{\alpha}_i \equiv$ estimador de α
- $\hat{\beta}_i \equiv$ estimador de β

Se estiman los parámetros poblacionales a partir de información muestral.

Estimador	$\rightarrow$	Fórmula
Estimado	$\rightarrow$	Valor numérico

FRM (forma estocástica): $y_i = \hat{\alpha} + \hat{\beta} x_i + \hat{u}_i$. Por lo tanto, $\hat{u}_i$ es el residual muestral (estimado de u_i).

2.5 Objetivo del análisis de regresión

Estimar FRP, $y_i = \alpha + \beta x_i + u_i$, a partir de estimar FRM, $y_i = \hat{\alpha} + \hat{\beta} x_i + \hat{u}_i$.

[IMAGEN]

2.6 Conceptos estadísticos

2.6.1 Sumatoria y multiplicatoria

$$\sum_{i=1}^{n} x_i = x_i + x_2 + ... + x_n$$

Propiedades del operador de sumatoria ($\sum$):

1. $\sum_{i=k}^{n} = nk \quad \therefore \quad k \equiv$ Constante.
2. $\sum_{i=1}^{n} kx_i = k \sum_{i=1}^{n} x_i \quad \therefore \quad k \equiv$ Constante.
3. $\sum_{i=1}^{n}(a + bx_i) = \sum_{i=1}^{n} a + \sum_{i=1}^{n} bx_i = na + b \sum_{i=1}^{n} x_i \quad \forall \quad a, b \equiv$ Constante.
4. $\sum_{i=1}^{n}(x_i + y_i) = \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i$.

Se pueden tener sumatorias múltiples:

$$\sum_{i=1}^{n}\sum_{j=1}^{m} = \sum_{i=1}^{n}(x_{i1} + x_{i2} + ... + x_{im})$$
$$= (x_{11} + x_{12} + ... + x_{1m}) + (x_{21} + x_{22} + ... + x_{2m}) + ... + (x_{n1} + x_{n2} + ... + x_{nm})$$

Propiedades del operador de sumatoria ($\sum\sum$):

1. $\sum_{i=1}^{n}\sum_{j=1}^{m} x_{ij} = \sum_{j=1}^{m}\sum_{i=1}^{n} \quad \rightarrow \quad$ Intercambiable
2. $\sum_{i=1}^{n}\sum_{j=1}^{m} x_i y_j = \sum_{i=1}^{n} x_i \sum_{j=1}^{m} y_j$
3. $\sum_{i=1}^{n}\sum_{j=1}^{m}(x_{ij} + y_{ij}) = \sum_{i=1}^{n}\sum_{j=1}^{m} x_{ij} + \sum_{i=1}^{n}\sum_{j=1}^{m} y_{ij}$

Adicionalmente, el operador multiplicatoria:

$$\Pi_{i=1}^{n} x_i = x_1 \cdot x_2 \cdot ... \cdot x_n$$

2.6.2 Valor esperado

El valor esperado de una variable aleatoria discreta se define como:

$$E(x) = \sum_{x} x f(x)$$

$$\begin{array}{ccc} \therefore \quad x & \rightarrow & \text{Valores de la variable aleatoria discreta} \\ f(x) & \rightarrow & \text{FDP (discreta) de x} \end{array}$$

En términos poblacionales:

$$E(x) = \mu_x \quad \rightarrow \quad \text{La media de la variable aleatoria discreta}$$

Propiedades del valor esperado $E(\cdot)$:

1. $E(b) = b \quad \therefore \quad b \equiv$ Constante
2. $E(ax + b) = E(ax) + E(b) = aE(x) + b \quad \forall \quad a, b \equiv$ Constantes
3. $E(xy) = E(x)E(y) \quad \forall \quad x, y \quad \rightarrow \quad$ Variables aleatorias independientes

2.6.3 Varianza

Sea x una variable aleatoria y $E(\cdot) = \mu$. La dispersión de valores de x alrededor de la media (valor esperado) es:

$$var(x) = \sigma_x^2 = E(x - \mu)^2$$

La desviación estándar de x es:

$$\text{Desv. est. }(x) = \sigma_x = [E(x-\mu)]^{1/2}$$

La varianza y desviación estándar muestral se calcula como:

$$s_x^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})}{n}, \quad s_x = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})}{n}}$$

Propiedades de la varianza:

1. $E(x-\mu) = E(x^2) - \mu^2 \quad \rightarrow \quad$ Nota: Al desarrollar el cuadrado se simplifica.
2. $var(c) = 0 \quad \because \quad c \equiv$ Constante
3. $var(ax+b) = var(ax) + var(b) = a^2 var(x) + 0 = a^2 var(x)$
4. $var(x+y) = var(x) + var(y) \quad \because \quad x, y$ son variables aleatorias independientes
5. $var(ax+by) = var(ax) + var(by) \quad \because \quad x, y$ son variables aleatorias independientes
6. $var(x \pm y) = var(x) + var(y) \pm 2cov(x,y) \quad \because \quad x, y$ son variables aleatorias correlacionadas

2.6.4 Covarianza

Sean x y y dos variables aleatorias con medias μ_x y μ_y, respectivamente. La covarianza se define como:

$$cov(x,y) = E[(x-\mu_x)(y-\mu_y)] = \sigma_{xy}$$

La covarianza muestral se calcula como:

$$cov(x,y) = \frac{\sum_x \sum_y (x_i-\bar{x})(y_i-\bar{y})}{n} = s_{xy}$$

Propiedades de la covarianza:

1. $cov(x,y) = E(xy) - \mu_x\mu_y = \mu_x\mu_y - \mu_x\mu_y = 0 \quad \because \quad x, y$ son independientes
2. $cov(a+bx, c+dy) = E\left[[(a+bx) - E(a+bx)]\,[(c+dy) - E(c+dy)]\right]$
   $= E\,[[bx - bE(x)]\,[dy - dE(y)]]$
   $= bd\,E\left[(x-\mu_x)(y-\mu_y)\right]$
   $= bd\,cov(x,y)$
3. $cov(c,x) = 0 \quad \because \quad c \equiv$ Constante
4. $cov(c,d) = 0 \quad \because \quad c, d$ son constantes

2.6.5 Coeficiente de correlación

El coeficiente de correlación (poblacional), se define como:

$$corr(x,y)=\rho=\frac{cov(x,y)}{\sqrt{var(x)var(y)}}=\frac{\sigma_{x,y}}{\sigma_x\sigma_y}$$

ρ es una medida de asociación lineal entre 2 variables $\rightarrow \quad -1\leq\rho\leq 1$. El coeficiente de correlación (muestral se tiene:)

$$corr(x,y)=\rho=\frac{\sum_x\sum_y(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_x(x_i-\bar{x})^2}\sqrt{\sum_y(y_i-\bar{y})^2}}$$

Capítulo 3

Modelo de regresión lineal simple

Suponer una FRP de dos variables:

$$y_i = \alpha + \beta x_i + u_i$$

Dado que la FRP no es observable, se estima la FRM:

$$\begin{array}{ccc} y_i = \hat{\alpha} + \hat{\beta} x_i + \hat{u}_i & \rightarrow & \hat{y}_i = \hat{\alpha} + \hat{\beta} x_i \therefore \\ y_i = \hat{y}_i + \hat{u}_i & & \because \hat{y}_i \equiv \text{Valor estimado de } \hat{y}_i \end{array}$$

¿Cómo se estima la FRM?

$$\begin{aligned} y_i &= \hat{y}_i + \hat{u}_i \therefore \\ \hat{u}_i &= y_i - \hat{y}_i \\ &= y_i - (\hat{\alpha} + \hat{\beta} x_i) \\ &= y_i - \hat{\alpha} - \hat{\beta} x_i \rightarrow \hat{u}_i \equiv \text{Residuos} \end{aligned}$$

3.1 Mínimos cuadrados ordinarios

Objetivo: Seleccionar la FRM de tal forma que la suma de los residuos al cuadrado sea la menor posible $\rightarrow \sum \hat{u}_i^2 = \sum (y_i - \hat{y}_i)^2$.

[IMAGEN]

Criterio de mínimos cuadrados:

$$
\begin{aligned}
\sum \hat{u}_i^2 &= \sum (y_i - \hat{y}_i)^2 \ \rightarrow \hat{y}_i = \hat{\alpha} + \hat{\beta} x_i \therefore \\
&= \sum (y_i - \hat{\alpha} - \hat{\beta} x_i)^2 \therefore \\
\min_{\hat{\alpha},\hat{\beta}} \sum \hat{u}_i^2 &= \sum (y_i - \hat{\alpha} - \hat{\beta} x_i)^2
\end{aligned}
$$

CPO (Condiciones de primer orden):

$$
\begin{aligned}
\frac{\partial \sum \hat{u}_i}{\hat{\partial} \alpha} &= 2 \sum (y_i - \hat{\alpha} - \hat{\beta} x_i)(-1) = 0 \\
&= -2 \sum (y_i - \hat{\alpha} - \hat{\beta} x_i) = 0 \\
&= \sum (y_i - \hat{\alpha} - \hat{\beta} x_i) = 0 \\
\sum y_i &= n\hat{\alpha} + \hat{\beta} \sum x_i \ \rightarrow \text{Ecuación normal (1)}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial \sum \hat{u}_i}{\hat{\partial} \beta} &= 2 \sum (y_i - \hat{\alpha} - \hat{\beta} x_i)(-x_i) = 0 \\
&= -2 \sum (y_i - \hat{\alpha} - \hat{\beta} x_i)(x_i) = 0 \\
&= \sum (y_i x_i - \hat{\alpha} x_i - \hat{\beta} x_i^2) = 0 \\
&= \sum y_i x_i - \hat{\alpha} \sum x_i - \hat{\beta} \sum x_i^2 = 0 \\
\sum y_i x_i &= \hat{\alpha} \sum x_i + \hat{\beta} \sum x_i^2 \ \rightarrow \text{Ecuación normal (2)}
\end{aligned}
$$

Retomemos las ecuaciones normales (1) y (2) y resolvemos para $\hat{\alpha}$ y $\hat{\beta}$:

$$
\left.\begin{aligned}
\sum y_i &= n\hat{\alpha} + \hat{\beta} \sum x_i \\
\sum y_i x_i &= \hat{\alpha} \sum x_i + \hat{\beta} \sum x_i^2
\end{aligned}\right\} \quad \text{Ax} = \text{d}
$$

$$
\text{A} = \begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix}_{2\times 2}, \quad \text{x} = \begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix}_{2\times 1}, \quad \text{d} = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix}_{2\times 1}
$$