Clase 1

Introducción a la Econometría

Profesor: Felipe Avilés Lucero

6 de octubre de 2009

Breve Definición e Historia

- Econometría es la ciencia que aplica métodos matemáticos y estadísticos al análisis de datos económicos, con el objetivo de dotar de una base empírica a una teoría económica, para así refutarla o verificarla.
- Aunque la econometría parece ser tan antigua como la misma ciencia económica, sólo en 1930 se crea la Sociedad Econométrica, la cual sistematizó su estudio y práctica.

Breve Definición e Historia

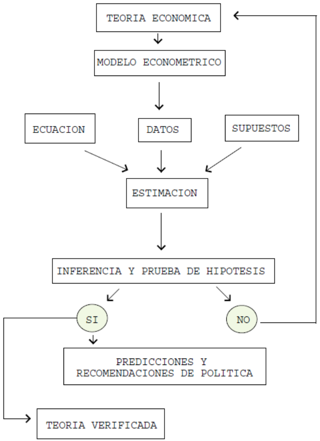
En 1933 se lanza el primer número de Econométrica en el que Ragnan Frish[1] destaca:

...la experiencia ha mostrado que cada uno de estos tres puntos de vista, el de la estadística, la teoría económica y las matemáticas, es necesario, pero por si mismo no suficiente para una comprensión real de las relaciones cuantitativas de la vida económica moderna. Es la unión de los tres aspectos lo que constituye una herramienta de análisis potente. Es la unión lo que constituye la econometría.

[1] Uno de los fundadores de la Sociedad Econométrica, a quién de hecho, se le acredita el haber acuñado el término Econometría.

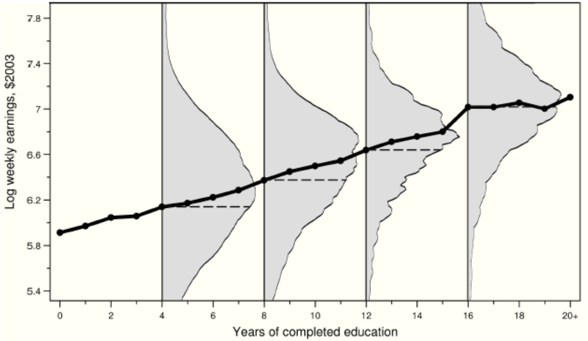
La Metodología Econométrica

La metodología econométrica se puede detallar de la siguiente forma:



De la Econonomía a la Estadística.

En el gráfico se muestran las medias de salario por nivel de escolaridad.



... para cada nivel de escolaridad, existe una distribución de salarios.

Relaciones Económicas y la Función de Esperanza Condicional (EC)

Dada la *aleatoriedad sistemática* en los resultados económicos, estos son difíciles de explicar... Sin embargo, esta *información* puede resumirse útilmente.

- En **promedio** personas con un mayor nivel de escolaridad obtienen mayores salarios que personas con menores años de educación. Carreras técnicas vs. universitarias.
- La conexión entre escolaridad y salarios posee un gran **poder predictivo**, a pesar de la gran heterogeneidad presente en los individuos. Por qué están Uds. aquí?

Cómo podemos resumir esta información de una manera convincente?
Función de Esperanza Condicional.

Relaciones Económicas y la Función de Esperanza Condicional (EC)

La función de EC de una variable dependiente y_i, dado un vector de regresores X_i de $K \times 1$ (con elemento x_{ki}) es la esperanza, o promedio poblacional de y_i cuando mantenemos X_i fijo. La denotaremos por $E[y_i|X_i]$.

Para un valor fijo de X_i, digamos $X_i = x$, escribimos $E[y_i|X_i = x]$.

Para una variable y_i continua, con función de densidad condicional $f_y(t|X_i = x)$ cuando $y_i = t$, la EC es:

$$E[y_i|X_i = x] = \int t f_y(t|X_i = x) dt$$

Relaciones Económicas y la Función de Esperanza Condicional (EC)

Un complemento importante de la función de EC es la ley de esperanzas iteradas:

$$E[y_i] = E\{E[y_i|X_i]\}$$

la demostración queda para Uds.

La ley de esperanzas iteradas posee una característica relevante, la de dividir una variable aleatoria en dos partes, la EC y un residuo con propiedades especiales que veremos a continuación.

Relaciones Económicas y la Función de Esperanza Condicional (EC)

Teorema 1: Propiedad de descomposición de la EC

$$y_i = E[y_i|X_i] + \epsilon_i$$

donde (i) ϵ_i es media independiente de X_i, esto es $E[\epsilon_i|X_i] = 0$, y por lo tanto (ii) ϵ_i no está correlacionado con ninguna función de X_i.

Demo.

(i) $E[\epsilon_i|X_i] = E[y_i - E[y_i|X_i]|X_i] = E[y_i|X_i] - E[y_i|X_i] = 0$.

(ii) Sea $h(X_i)$ cualquier función de X_i. Por la ley de esperanzas iteradas, $E[h(X_i)\epsilon_i] = E\{h(X_i)E[\epsilon_i|X_i]\}$, y por independencia en media, $E[\epsilon_i|X_i] = 0$.

Relaciones Económicas y la Función de Esperanza Condicional (EC)

Teorema 2: Propiedad predictiva de la función de EC.
Sea $m(X_i)$ cualquier función de X_i.
La función de EC resuelve

$$E(y_i|X_i) = \text{argmin}_{m(X_i)} E[(y_i - m(X_i))^2]$$

por lo tanto, la función de EC es el predictor de mínimo error cuadrático medio de y_i dado X_i.
Demo.

$$\begin{aligned}(y_i - m(X_i))^2 &= ((y_i - E(y_i|X_i)) + (E(y_i|X_i) - m(X_i))^2 \\ &= (y_i - E(y_i|X_i))^2 + 2(E(y_i|X_i) - m(X_i)) \\ &\times (y_i - E(y_i|X_i)) + (E(y_i|X_i) - m(X_i))^2\end{aligned}$$

Relaciones Económicas y la Función de Esperanza Condicional (EC)

Cont. Demo.

$$
\begin{aligned}
(y_i - m(X_i))^2 &= (y_i - E(y_i|X_i))^2 + 2(E(y_i|X_i) - m(X_i)) \\
&\times (y_i - E(y_i|X_i)) + (E(y_i|X_i) - m(X_i))^2
\end{aligned}
$$

El primer término de la ecuación anterior no involucra $m(X_i)$. El segundo puede escribirse como $h(X_i)\epsilon_i$, donde $h(X_i) \equiv 2(E(y_i|X_i) - m(X_i))$ que posee esperanza igual a 0 (por teorema 1). El último término se minimiza en 0 cuando $m(X_i)$ es la función de EC.

Definición de Regresión

La regresión es un elemento fundamental en la Econometría, corresponde a un estudio de dependencia entre una variable dependiente y una o más variables explicativas.

El análisis de regresión tiene como objeto estimar y/o predecir el promedio poblacional de la variable dependiente para valores fijos de la(s) variable(s) explicativa(s).

Análisis de Regresión Poblacional

Anteriormente vimos que la función de EC $E(Y|X_i)$ es función de X_i, donde X_i es un valor dado de X:

$$E(Y|X_i) = f(X_i)$$

donde $f(X_i)$ es una función cualquiera. La ecuación anterior se denomina Regresión Poblacional. Que forma tiene $f(X_i)$ es una pregunta empírica, aunque muchas veces la teoría nos puede ayudar bastante.

Análisis de Regresión Poblacional

Supongamos que en nuestro ejemplo anterior el salario esta relacionado linealmente con la educación, así podemos suponer que la función de regresión poblacional $E(y_i|X_i)$ es una función lineal de X_i, es decir:

$$E(y_i|X_i) = \beta_1 + \beta_2 X_i$$

donde β_1 y β_2 se denominan coeficientes de regresión. Así el objetivo es estimar β_1 y β_2 a partir de datos de X e y. Por lo tanto podemos reescribir y_i de la siguiente forma:

$$\begin{aligned} y_i &= E(y_i|X_i) + \epsilon_i && \text{(por Teo. 1)} \\ y_i &= \beta_1 + \beta_2 X_i + \epsilon_i \end{aligned}$$

Análisis de Regresión Muestral

En la mayoría de los fenómenos económicos a estudiar, no disponemos de las observaciones totales de la **población** (ej. CENSO).

En la práctica se tiene alcance nada más que a una **muestra** de los valores de y que corresponden a unos valores fijos de X (ej. CASEN). En este caso tenemos que estimar la función de regresión poblacional en base a información muestral.

Análisis de Regresión Muestral

Supongamos que los datos poblaciones de ingreso disponible (YD) y consumo (C) para cierta población objetivo son los siguientes:

Y|X	80	100	120	140	160	180	200	220	240	260
Gasto en	55	65	79	80	102	110	120	135	137	150
consumo	60	70	84	93	107	115	136	137	145	152
familiar	65	74	90	95	110	120	140	140	155	175
semanal	70	80	94	103	116	130	144	152	165	178
(Y)	75	85	98	108	118	135	145	157	175	180
	-	88	-	113	125	140	-	160	189	185
	-	-	-	115	-	-	-	162	-	191
Media Condicional	65	77	89	101	113	125	137	149	161	173

Supongamos que nosotros no conocemos estos datos, es decir, no tenemos acceso a las observaciones correspondientes a la población total.

Análisis de Regresión Muestral

Tenemos a nuestra disposición sólo una muestra, la que ha sido obtenida de forma aleatoria de la población.

Y	X
70	80
65	100
90	120
95	140
110	160
115	180
120	200
140	220
155	240
150	260

Y	X
55	80
88	100
90	120
80	140
118	160
120	180
145	200
135	220
145	240
175	260

Es importante notar que a partir de una población podemos sacar una gran cantidad de muestras en forma aleatoria y en la realidad nosotros observamos sólo una de ellas (uno de los dos cuadros).

Análisis de Regresión Muestral

Entonces es importante tener en mente que las rectas de regresión muestral representan la recta de regresión poblacional, pero debido a fluctuaciones muestrales pueden ser consideradas sólo como una aproximación.

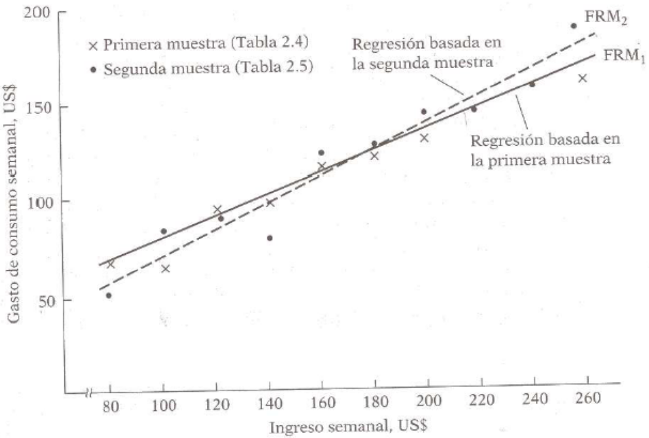
Como contraparte muestral la función de regresión muestral puede escribirse como:

$$\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$$

donde $\hat{y}_i$ es el estimador de $E(y_i|X_i)$, $\hat{\beta}_1$ es el estimador de β_1 y $\hat{\beta}_2$ es el estimador de β_2.

Análisis de Regresión Muestral

En el grafo vemos rectas de regresión en base a muestras distintas



Análisis de Regresión Muestral

Definición: Un *estimador* es una regla, fórmula o método que dice cómo determinar el parámetro poblacional a partir de la información suministrada por la muestra disponible.

De igual manera que para el caso poblacional la función de regresión muestral también tiene una representación estocástica:

$$y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{\epsilon}_i$$

Entonces, el objetivo del Análisis de Regresión es estimar la Función de regresión poblacional:

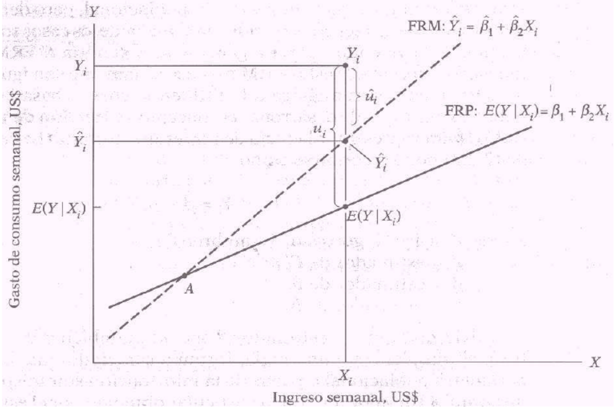
$$y_i = \beta_1 + \beta_2 X_i + \epsilon_i$$

con base en la Función de regresión muestral:

$$y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{\epsilon}_i$$

Análisis de Regresión Muestral

Rectas de regresión muestral y poblacional



Análisis de Regresión Muestral

En términos de la función de regresión muestral, la y_i observada puede ser expresada como:

$$y_i + \hat{y}_i + \hat{\epsilon}_i$$

y en términos de la función de regresión poblacional puede ser expresada como:

$$y_i = E(y_i|X_i) + \epsilon_i$$

En el cuadro anterior podemos notar que para todo X_i a la derecha del punto A, $\hat{y}_i$ sobreestima $E(Y|X_i)$. De igual manera, para cualquier punto a la izquierda de A, $\hat{y}_i$ subestima $E(Y|X_i)$. Esta sobreestimación y subestimación del modelo poblacional es inevitable debido a las fluctuaciones muestrales.

Propiedades de un estimador

- Se denomina *sesgo* a la diferencia entre el valor esperado del estimador y su verdadero valor: $E(\hat{\beta}) - \beta$. De esta forma, se dice que $\hat{\beta}$ es un estimador insesgado si $E(\hat{\beta}) = \beta$.
- El estimador es *eficiente* o de mínima varianza si no hay ningún otro estimador insesgado que tenga una varianza menor que $\hat{\beta}$. En general se trata de utilizar estimadores de varianza pequeña, pues de este modo la estimación es más precisa.
- El *Error Cuadrático Medio* (ECM) es una propiedad de los estimadores que mezcla los conceptos de eficiencia e insesgamiento. El ECM de $\hat{\beta}$ se define como:

$$ECM(\hat{\beta}) = E[(E(\hat{\beta}) - \beta)^2]$$

Lo que se puede expresar equivalentemente de la siguiente manera: