ANÁLISIS DE REGRESIÓN CON DOS VARIABLES: ALGUNAS IDEAS BÁSICAS

El análisis de regresión más sencillo se da con el caso de dos variables, conocido como la ***regresión bivariante o de dos variables*** (regresión simple). En esta regresión la variable explicada se relaciona con una sola variable explicativa.

En el capítulo anterior se llegó a la conclusión que hoy en día el análisis de regresión trata de:

EL ESTUDIO DE LA DEPENDENCIA DE LA VARIABLE DEPENDIENTE, RESPECTO A UNA O MÁS VARIABLES (EXPLICATIVAS), CON EL OBJETIVO DE ESTIMAR Y/O PREDECIR LA MEDIA O VALOR PROMEDIO POBLACIONAL DE LA PRIMERA EN TÉRMINOS DE LOS VALORES CONOCIDOS O FIJOS (EN REPETIDAS MUESTRAS) DE LAS ÚLTIMAS.

Esta afirmación se transforma en: $E(Y|X)$

Se quiere estimar el valor promedio de Y en función de los valores fijos de X. Esto significa ver cómo cambia Y cuando cambia X:

$$Y = f(X)$$

Si en los modelos econométricos las relaciones no son exactas:

¿CÓMO SE TOMAN EN CUENTA LOS OTROS FACTORES QUE INFLUYEN SOBRE Y?; ¿CÚAL ES LA RELACIÓN FUNCIONAL ENTRE Y y X?; ¿CÓMO SE CAPTA LA RELACIÓN CETERIS PARIBUS ENTRE Y y X?.

$$Y = \beta_1 + \beta_2 X + \varepsilon$$

Valor promedio de Y que no depende de X.

$$\frac{\Delta Y}{\Delta X}$$

TÉRMINO DE ERROR ALEATORIO O PERTURBACIÓN ESTOCÁSTICA: conocido como COMPONENTE ALEATORIO agrupa todo los factores que afectan a Y que no son incluidos en el modelo.

FUNCIÓN DE REGRESION POBLACIONAL (FRP): denota que el valor esperado de la distribución de Y dada X, está relacionada funcionalmente con X. También se le denomina ***función de expectativa condicional (FEC) o regresión poblacional (RP).***

$$E(Y|X_i) = f(X_i) \longrightarrow \text{FRP}$$

$$\downarrow$$

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

Coeficientes de regresión. $\boldsymbol{\beta_1}$ es el intercepto y $\boldsymbol{\beta_2}$ se conoce como el coeficiente de pendiente. Parámetros no conocidos pero fijos.

¿SE PUEDE TRABAJAR CON LA FRP?

ESPECIFICACIÓN ESTOCÁSTICA DE LA FRP

$$\mu_i = Y_i - E(Y|X_i)$$

$$FRP \qquad Y_i = E(Y|X_i) + \varepsilon_i$$

Componente sistemático

Componente aleatorio

ε_i es una variable aleatoria no observable que toma valores negativos o positivos. Se conoce como ***perturbación estocástica o término de error estocástico.***

$$\begin{aligned} Y_i &= E(Y|X_i) + \varepsilon_i \\ &= \beta_1 + \beta_2 X + \varepsilon_i \end{aligned}$$

Tomando el valor esperado de Y_i

$$\begin{aligned} E(Y_i|X_i) &= E[E(Y|X_i)] + E(\varepsilon_i|X_i) \\ &= E(Y|X_i) + E(\varepsilon_i|X_i) \end{aligned}$$

Dado que $E(Y_i|X_i) = E(Y|X_i)$ entonces: $E(\varepsilon_i|X_i) = 0$

SIGNIFICADO DEL TÉRMINO “PERTURBACIÓN ESTOCÁSTICA”

1. Vaguedad de la teoría.
2. No disponibilidad de información.
3. Variables centrales (relevantes) Vs variables periféricas (secundarias)
4. Aleatoriedad intrínseca en el comportamiento humano.
5. Variables proxy (errores de medida)
6. Principio de parsimonia.
7. Forma funcional incorrecta.

FUNCIÓN DE REGRESIÓN MUESTRAL (FRM)

En la práctica lo que se tiene a la mano es información muestral y no poblacional, por ende la FRP es desconocida.

¿QUÉ HACER?

R: ESTIMAR LA FRP CON BASE EN INFORMACIÓN MUESTRAL[1].

$$FRM \qquad \hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$$

$\hat{Y}_i$: estimador de $E(Y|X_i)$

$\hat{\beta}_1$: estimador de β_1

$\hat{\beta}_2$: estimador de β_2

ESTIMADOR (estadístico): es una regla, fórmula o método que dice cómo estimar el parámetro poblacional a partir de información muestral.

ESTIMACIÓN: valor numérico particular obtenido por el estimador en una aplicación.

[1] En general se obtendrán N FRM diferentes para N muestras diferentes y estas FRM no necesariamente serán iguales.

Forma estocástica de la FRM: $Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{\varepsilon}_i$

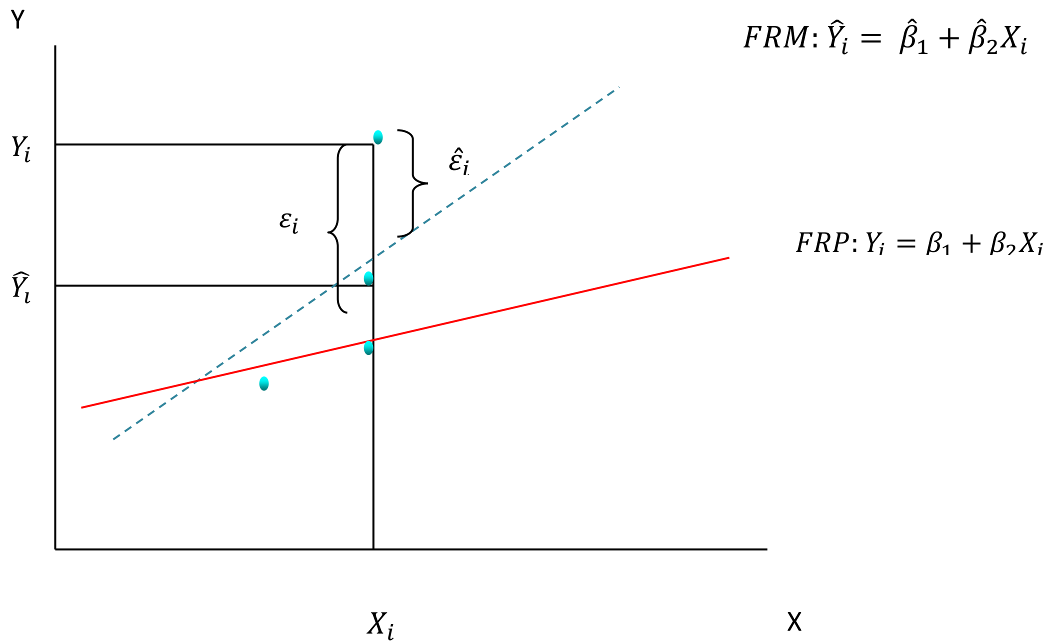
EL OBJETIVO PRINCIPAL EN EL ANÁLISIS DE REGRESIÓN ES ESTIMAR LA FRP CON BASE EN LA FRM

$$Y_i = \beta_1 + \beta_2 X + \hat{\varepsilon}_i$$

$$\Big\uparrow \quad \text{Inferencia}$$

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

$$Y_i = \hat{Y}_i + \hat{\varepsilon}_i$$



LINEALIDAD: UN ASUNTO PARA TENER PRESENTE

Cuando se habló de la clasificación de los modelos econométricos se explicó que durante este curso solamente se emplearían modelos lineales. Existen dos tipos de linealidades:

1) ***LINEALIDAD EN LAS VARIABLES:*** el significado más simple de linealidad es aquel que indica que la esperanza condicional de Y es una función lineal de X_i. Geométricamente la curva de regresión en una línea recta.

$E(Y|X_i) = \beta_1 + \beta_2 X_i$ $\longrightarrow$ *LINEAL EN VARIABLES*

$E(Y|X_i) = \beta_1 + \beta_2 {X_i}^2$ $\longrightarrow$ *NO LINEAL EN VARIABLES.*

2) LINEALIDAD EN PARÁMETROS: ocurre cuando la esperanza condicional de Y, $E(Y|X_i)$ es una función lineal de los parámetros β, independientemente de la linealidad en las variables.

LINEALIDAD EN PARÁMETROS

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

$$E(Y|X_i) = \beta_1 + \beta_2 {X_i}^2$$

La linealidad relevante durante el curso es la linealidad en parámetros para poder desarrollar la teoría de regresión.

$$Y = e^{\beta_1 + \beta_2 X}$$

$$Y = \beta_1 + \beta_2 X + \beta_3 X^2$$

$$Y = \beta_1 + \beta_2 X + \beta_3 X^2 + \beta_4 X^3$$

NO SE OLVIDE QUE CUANDO SE HABLA DE REGRESION LINEAL SIEMPRE SIGNIFICARÁ QUE ES LINEAL EN PARÁMETROS, ES DECIR LOS β ELEVADOS A LA PRIMERA POTENCIA.

"Mientras que la mecánica de la regresión simple no depende de la manera en que estén definidas X e Y, la interpretación de los coeficientes sí depende de sus definiciones. Para realizar un trabajo empírico exitoso es mucho más importante la capacidad de interpretar los coeficientes que tener destrezas para calcularlos" Wooldridge 2010.

Lecturas obligatorias:

- ✓ Gujarati, D. y Porter, D. (2010). Econometría. 5ta. Edición McGraw Hill. Capítulo 2.
- ✓ Novales, A. (1993). Econometría. 2da. Edición. McGraw Hill. Interamericana. Capítulo 3.