Econometria
CEF y Proyección Lineal

Pasquini, Ricardo

Facultad Ciencias Empresariales Austral
Universidad Austral

March 12, 2023

CEF y Proyección Lineal

- Motivación: Descripción del ingreso en la población
- Funcion de Esperanza Condicional (CEF) y sus propiedades
- Varianza Condicional y Varianza del Error

Ejemplo

Analizaremos la teoría junto al caso de los ingresos individuales en CABA.

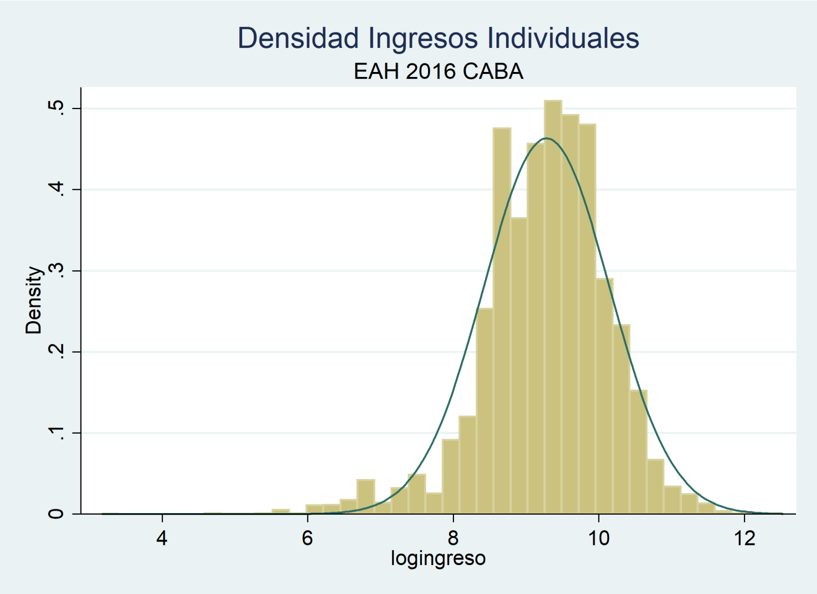
Distribuciones Poblacionales

- Supondremos Y proveniente de una población con una CDF

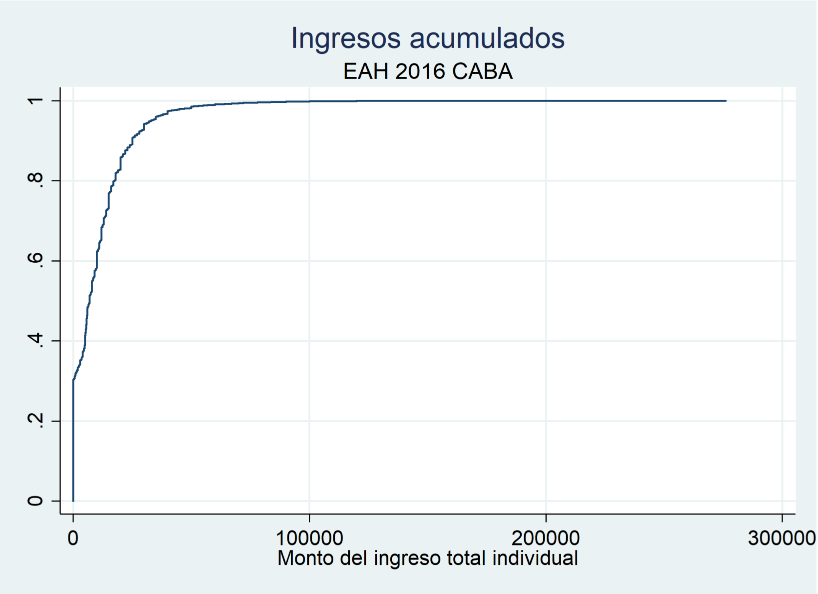
$$F(y) = Prob(Y \leq y)$$

- Supondremos factores explicativos como $X_1, X_2, \ldots, X_k$ tambien como variables aleatorias con sus respectivas distribuciones.

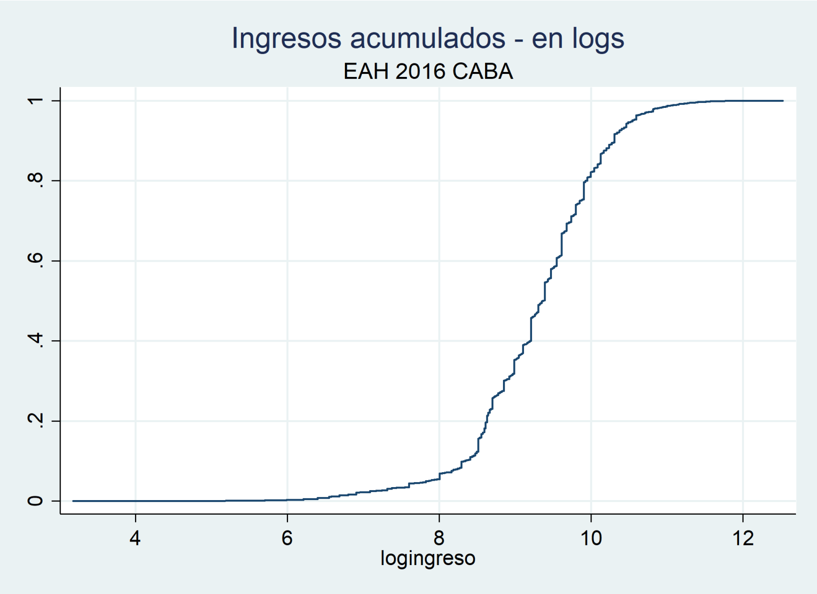
Distribución del ingreso - Densidad



Distribución del ingreso - Densidad Acumulada



Distribución del ingreso- Densidad Acumulada

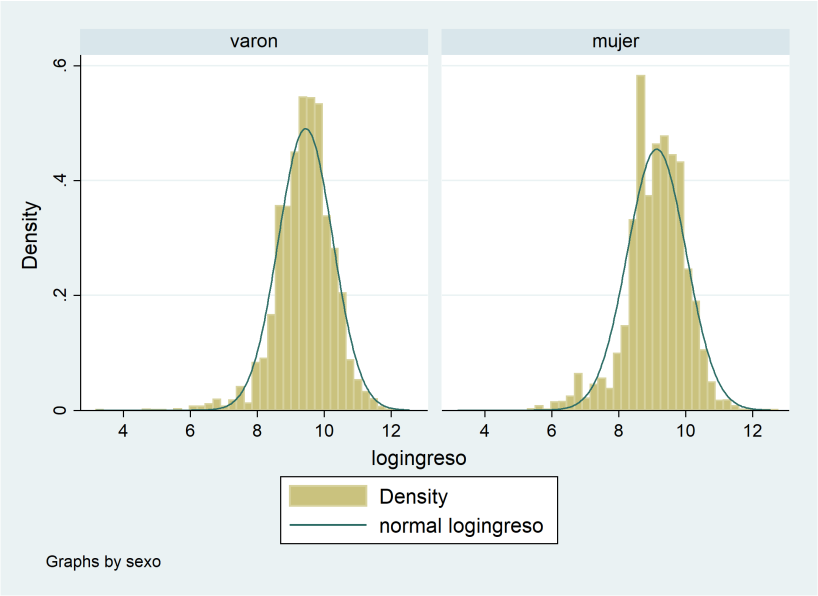


Distribucion del ingreso por sexo

¿Explica el sexo la distribución del ingreso? ¿Varía la distribución del ingreso de acuerdo al sexo?

- Lo inspeccionaremos gráficamente
- Utilizaremos el valor esperado condicional como una aproximación

Distribucion del ingreso por sexo



Distribucion del ingreso por sexo

Aproximación: Esperanza Condicional

$$E[Y|sexo = "hombre"] = 9.44$$

$$E[Y|sexo = "mujer"] = 9.13$$

```
. mean logingreso, over(sexo)

Mean estimation                   Number of obs    =     10,113

       varon: sexo = varon
       mujer: sexo = mujer

--------------------------------------------------------------
        Over |       Mean   Std. Err.     [95% Conf. Interval]
-------------+------------------------------------------------
logingreso   |
       varon |   9.444136   .0117493      9.421105    9.467167
       mujer |   9.136385   .0120212      9.112822    9.159949
--------------------------------------------------------------
```

Función de Esperanza Condicional (CEF)

- En general, es natural que para un valor de x estemos interesados en conocer el valor esperado. Lo definimos como:

$$E[Y|x] \equiv m(x)$$

- Notar que puesto que x es una variable aleatoria entonces $m(x)$ también es una variable aleatoria.

Función de Esperanza Condicional (CEF)

- Notemos que un modelo simple para explicar o predecir sería:

$$Y = m(x) + e$$

donde

$$e \equiv Y - m(x)$$

- Algunas propiedades del error e:
  1. $E[e|x] = 0$
  2. $E[e] = 0$

Función de Esperanza Condicional (CEF)

Propiedades del error

- $E[e|x] = 0$

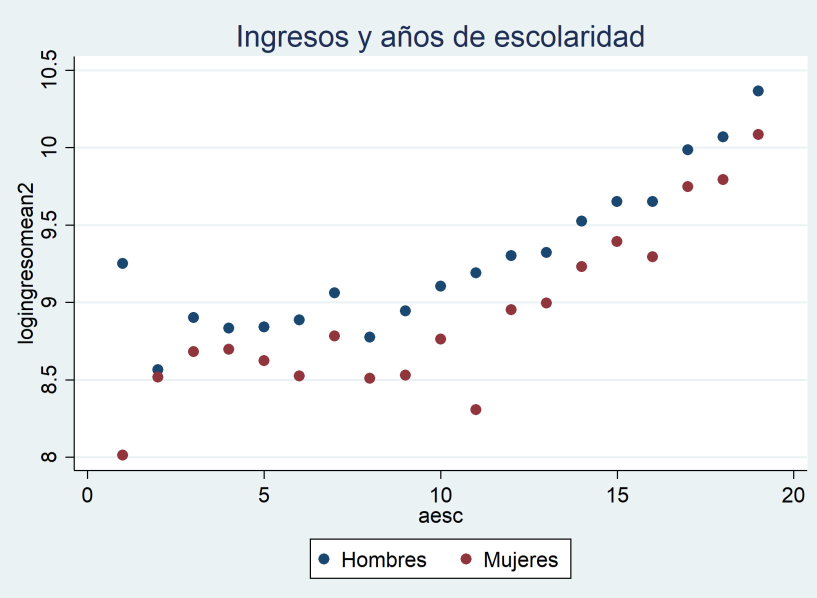
  $$E[e|x] = E[y - m(x)|x] = E[y|x] - E[m(x)|x] = m(x) - m(x) = 0$$

- $E[e] = 0$

  $$E[e] = E[E[e|x]] = E[0] = 0$$

Funcion de Esperanza Condicional

Aplicacion: Ingresos y Años de Escolaridad



Problema de Predicción

- La función CEF tiene una propiedad teórica interesante: provee la mejor predicción en un sentido específico.
- Supongamos que dado un vector de caracteristicas x queremos buscar una funcion $g(x)$ que nos haga la mejor predicción posible sobre y. Una forma de definir mejor predicción, es pedir que minimice el error cuadrático esperado

$$E[(y - g(x))^2]$$

Problema de Predicción

$g(x) = E[y|x] \equiv m(x)$ como la solución

- Se puede demostrar que la función que minimiza el error cuadrático medio es $g(x) = E[y|x]$.

Proof.

$$E[(y - g(x))^2] = E[(e + m(x) - g(x))^2]$$

$$= \mathrm{E}(\mathrm{e}^2) + 2E(e(m(x) - g(x))) + E((m(x) - g(x))^2)$$

$$= \mathrm{E}(\mathrm{e}^2) + E((m(x) - g(x))^2) > E(e^2) = E((y - m(x))^2) \quad \square$$

Problema de Predicción

$g(x) = E[y|x] \equiv m(x)$ como la solución

- Una desventaja es que no siempre será fácil estimar $E[y|x]$, por ejemplo por tener pocos datos para nuestro x de interés.
- Tampoco conocemos la forma funcional de $E[y|x]$

Varianza Condicional

Nuestro objeto de interés es más que el valor esperado

- Definimos varianza condicional en general como:

$$Var(w|x) = E[(w - E[w|x])^2]$$

- Se sigue que la *varianza condicional del error del modelo CEF* es la esperanza condicional del error al cuadrado:

$$\sigma^2(x) = Var(e|x) = E[(e - E[e])^2] = E[e^2|x]$$

Y definimos tambien el *desvío estandar condicional*:

$$\sigma(x) = \sqrt{E[e^2|x]}$$

Notar que la varianza del error *no-condicional* es el valor esperado de la varianza condicional

$$\sigma^2 = E[e^2] = E[E[e^2|x]] = E(\sigma^2(x))$$