Econometria
CEF y Proyección Lineal

Pasquini, Ricardo

Facultad Ciencias Empresariales Austral
Universidad Austral

March 1, 2024

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 es una *variable aleatoria* proveniente de una población con una función de densidad acumulativa (CDF)

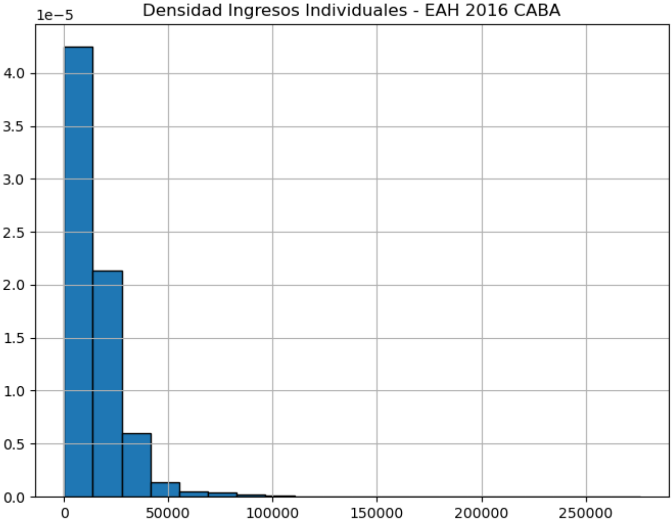
$$F(y) = Prob(Y \leq y)$$

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

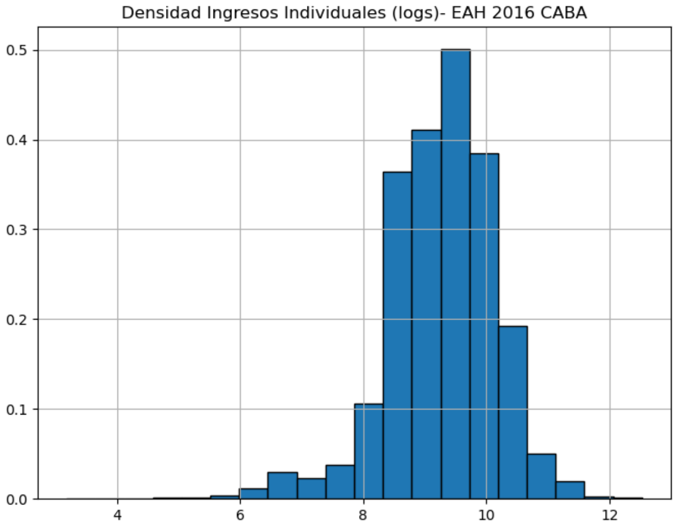
Repaso Estadístico

- Una *variable aleatoria* es una variable cuyo resultado es a priori desconocido y queda determinada por un experimento
- En este caso las variables provienen de la distribución de la población.

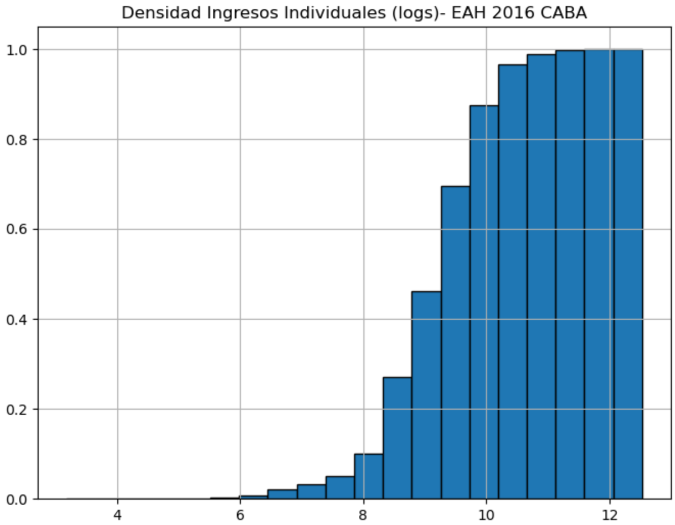
Distribución del ingreso - Densidad



Distribución del ingreso - Densidad



Distribución del ingreso- Densidad Acumulada

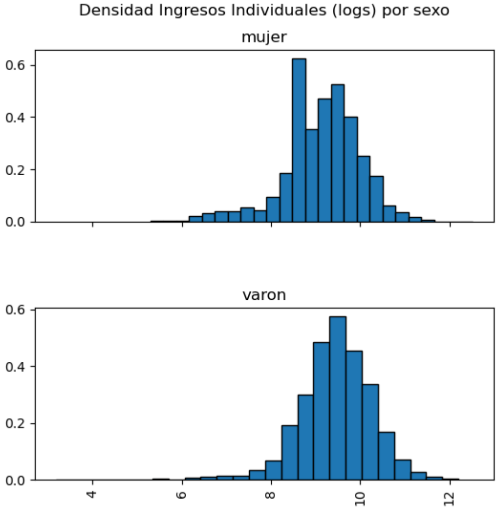


Distribucion del ingreso por sexo

¿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: Promedio muestral

- Intuitivamente un valor que sirve para describir la muestra es el promedio (Avg)

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

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

	count	mean	std	min	25%	50%	75%	max
sexo								
mujer	5324.000000	9.136385	0.877134	5.298317	8.612503	9.210340	9.718660	12.528156
varon	4789.000000	9.444136	0.813080	3.178054	8.987197	9.472705	9.928180	12.206073

Table: Descripcion de Ingresos

Distribucion del ingreso por sexo

Aproximación: Esperanza Condicional

- Nuestro objetivo no es solo describir una muestra. Queremos establecer una teoría a nivel poblacional.
- Proponemos la esperanza matemática

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

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

Repaso Estadístico

- Si X es una variable aleatoria, el valor esperado (o expectativa) de X, denotado $E[X]$ y a veces μ_X o simplemente μ, es un promedio ponderado de todos los valores posibles de X. Los pesos están determinados por la función de densidad de probabilidad.
- Suponiendo un caso discreto donde X toma k posibles valores

$$E[X] = \sum_{j=1}^{k} X_k P(X_k)$$

Repaso Estadístico

- Suponiendo un caso continuo tenemos

$$E[X] = \int_{-\infty}^{\infty} Xf(X)dx$$

Repaso Estadístico

Propiedades de la Esperanza

1. Para cualquier constante

$$E[c] = c$$

2. Para cualquier constante a y b

$$E[aX + b] = aE[X] + b$$

3. Si $\{a_1, a_2..a_k\}$ y $\{X_1, X_2..X_k\}$

$$E[\sum a_i X_i] = \sum a_i E[X_i]$$

Función de Esperanza Condicional (CEF)

- En general, es natural que estemos interesados en el ingreso *de las mujeres* , de los hombres, es decir para un subgrupo.
- En terminos estadísticos es natural que estemos interesados en conocer el valor esperado de una variable Y condicional a un cierto valor de $X = x$. Lo definimos como:

$$E[Y|x] = \sum y_k f_{Y|X}(y_k|x) \equiv m(x)$$

- 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

- $E[e|x] = 0$

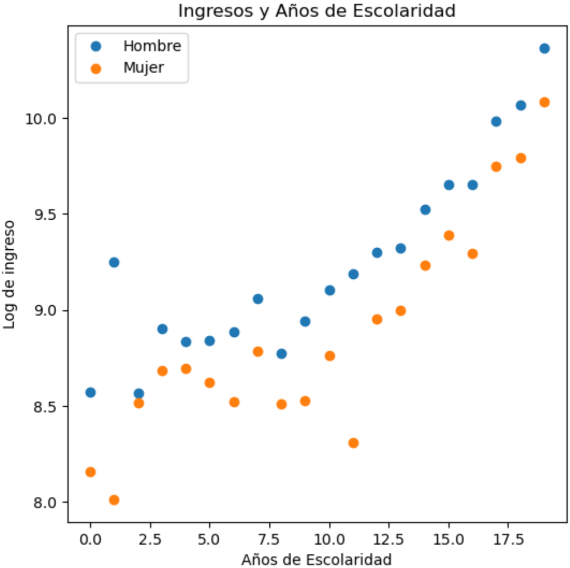
  $$E[e|x] = E[y - m(x)|x] = E[y|x] - E[m(x)|x] = m(x) - m(x) = 0$$

- $E[e] = 0$

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

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

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