

## Regresión por Cuantiles

Autor: Univ. Deyvis Nina Canaviri

### 1. Esperanzas condicionales y Regresión.

Consideremos un modelo lineal simple:

$$y = x' \beta + u$$

con  $E(u|x) = 0$ . De modo que:

$$E(y|x) = x' \beta$$

$x' \beta$  es una función de regresión, es la esperanza condicional de  $y$  en  $x$ .

Trivialmente:

$$\frac{\partial E(y|x)}{\partial x_k} = \beta_k$$

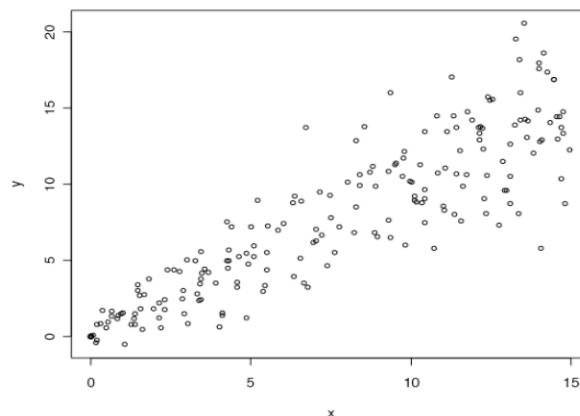
$\beta$  mide como cambios marginales en  $x$  afectan a  $E(y|x)$ .

Si  $u$  es independiente de  $x$ , también es cierto que:

$$\frac{\partial y}{\partial x} = \beta$$

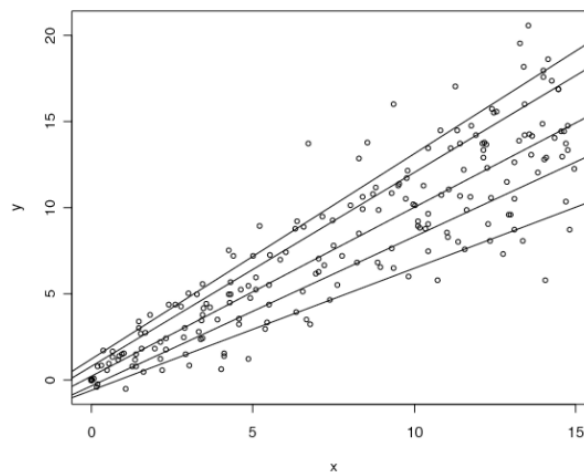
- La independencia es lo que le da sentido a la idea de “alterar  $x$  dejando  $u$  constante”.
- El efecto de  $x$  sobre  $E(y|x)$  resume también el efecto de  $x$  sobre  $y$ .

Considerando el siguiente caso de *heterocedasticidad*:



❖ ¿Cuál es el efecto de  $x$  sobre  $E(y|x)$ ?

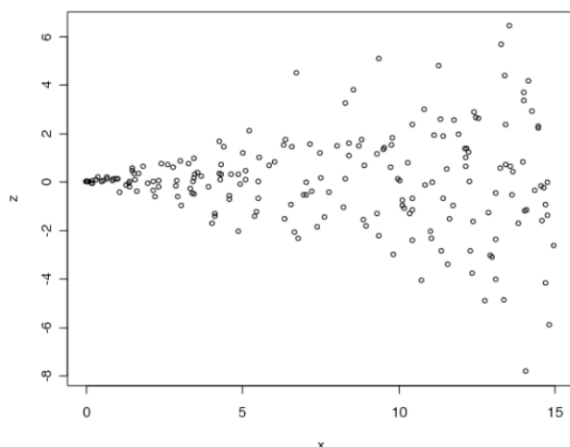
❖ ¿Cuál es el efecto de  $x$  sobre  $y$ ?



✓ El efecto de  $x$  es mayor *arriba*. El efecto no es *homogéneo*.

✓  $\beta = \partial E(y|x) / \partial x$  no resume el ejemplo de  $x$  sobre  $y$ .

❖ ¿Es posible que  $x$  no afecte a  $E(y|x)$  pero que sí a  $y$ ?



$x$  No afecta a  $E(y|x)$  pero si afecta a  $y$ .

## 2. Regresión por cuantiles.

Intenta modelar el efecto de  $x$  sobre toda la distribución de  $y$ .

$Z \sim F(z)$  continua y monótona.

El  $\tau$ -esimo cuantil de  $Z$  es un número  $Q_Z(\tau)$  que satisface:

$$F(Q_Z(\tau)) = \tau$$

O sea, el  $\tau$ -esimo cuantil es un número de la distribución tal que la probabilidad de que ocurran valores menores es  $\tau$ .

## 3. Cuantiles de la distribución normal estándar

Recordar que el modelo simple de regresión puede ser visto como:

$$E(y|x) = x' \beta$$

En forma analoga, el modelo de regresión para el  $\tau$ -esimo cuantil de la distribución de  $y$  condicional en  $x$  sera:

$$Q_{y|x}(\tau) = x' \beta(\tau)$$

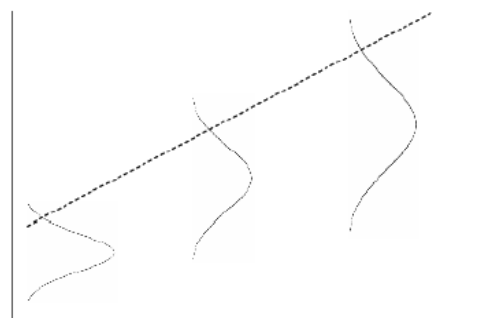
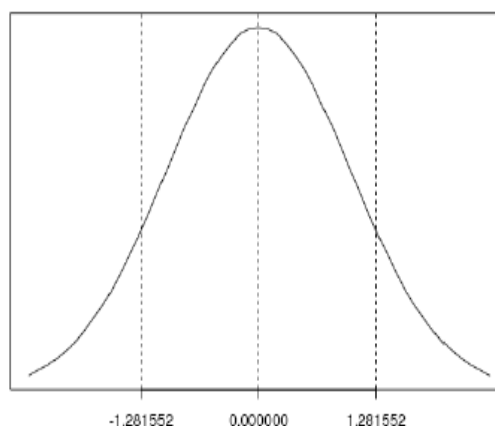
Con:

$$\frac{\partial Q_{y|x}(\tau)}{\partial x} = \beta(\tau)$$

Estamos permitiendo que el efecto de  $x$  sobre  $y$  sea distinto en distintos lugares de la distribución de  $y$  dado  $x$ .

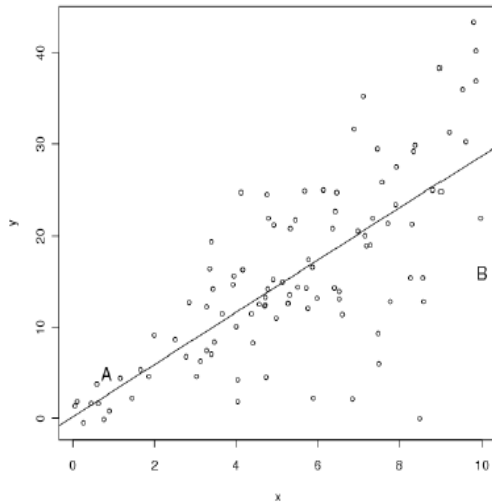
Ejemplo:

$$Q_{y|x}(0.75) = \beta_0(0.75) + \beta_1(0.75)x$$



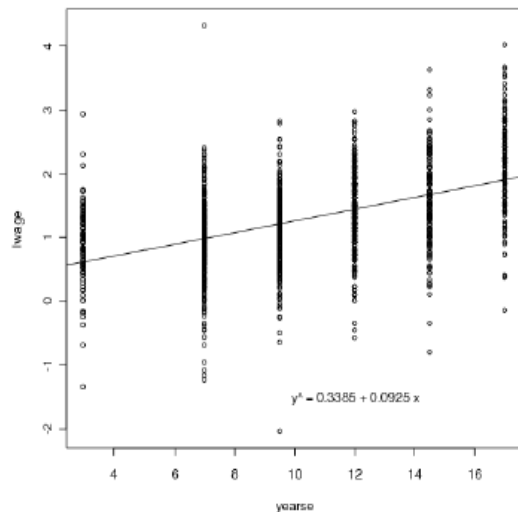
La recta une los Cuantiles 0.75 de cada distribución condicional.

Importante: el método estima rectas para distintos lugares de la distribución condicional (y no de la no condicional!!!!)



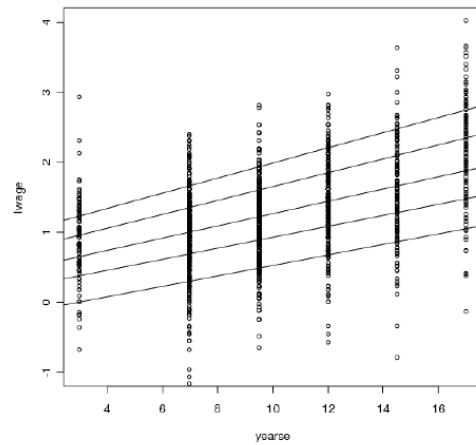
- “A” está arriba de la distribución condicional y abajo en la no condicional.
- “B” está en el medio de la no condicional, y muy abajo en la condicional.

Ejemplo: Retorno a la Educación



- El retorno “medio” es 0.0925

Estimación por quantile regression:



- El retorno es superior en los niveles más altos.

Resumen de resultados:

| Cuantiles           | intercepto | educacion |
|---------------------|------------|-----------|
| 0.10                | -0.2260    | 0.0753    |
| 0.25                | 0.1420     | 0.0787    |
| 0.50                | 0.3841     | 0.0881    |
| 0.75                | 0.6580     | 0.0992    |
| 0.90                | 0.9055     | 0.1083    |
| Retorno medio (MCO) | 0.3385     | 0.0925    |

Estimación e Inferencia

Mínimos cuadrados ordinarios:

$$\text{armin} \sum (y_i - \hat{\alpha} - \hat{\beta}x_i)^2$$

- La solución es una recta en (X,Y).
- La recta pasa por el medio (los errores son penalizados en forma simétrica).

Koenker y Bassett (1978): la solución al problema pasa por penalizar Asimétricamente los errores de estimación.

$$\hat{\beta}(\tau) = \text{armin} \sum_{i=1}^n \rho_{\tau}(y_i - x_i\beta)$$

Con  $\rho_{\tau}(Z) = z(\tau - I(z < 0))$ ,  $\tau \in (0,1)$ , produce estimadores consistentes y asintóticamente normales.

- Penaliza a los errores positivos con  $\tau$  y a los negativos con  $1 - \tau$

- Cuando  $\tau = 0.5$ , mínimos desvíos absolutos, penalización simétrica.

El problema de estimación puede ser reescrito como:

$$\min_{(b(\tau), u, v) \in \mathbb{R} \times \mathbb{R}_+^{2n}} \{\tau 1_n' u + (1 - \tau) 1_n' v \mid 1_n' b(\tau) + u - v = y\}$$

$e_i \equiv x_i' b(\tau)$ ,  $1_n$  un vector columna de  $n$ ,  $u$  y  $v$  son variables de holgura complementaria: es un Programa lineal.

En la Practica estimamos  $\hat{\beta}(\tau_i), i = 1, \dots, m$  (coeficientes para  $m$  cuantiles)

Denominemos con  $\beta(\tau_i)$  a los coeficientes poblacionales y construyamos los siguientes vectores.

- $\hat{\beta} \equiv (\hat{\beta}(\tau_1)' \cdots \hat{\beta}(\tau_m)')'$
- $\beta \equiv (\beta(\tau_1)' \cdots \beta(\tau_m)')'$

## 4. Inferencia

Bajo el supuesto de que la muestra es independiente (pero no necesariamente idénticamente distribuida) y bajo condiciones de regularidad estándar, es posible mostrar que:

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(0, \Lambda)$$

$$\Lambda = \Lambda_{j,p}, j = 1, \dots, m, p = 1, \dots, m \text{ con:}$$

$$\Lambda_{j,p} = (\min\{\tau_j, \tau_p\} - \tau_j \tau_p) \left( E[f_{\tau_j}(0|x) xx'] \right)^{-1} E[xx'] \left( E[f_{\tau_j}(0|x) xx'] \right)^{-1}$$

en donde  $f_{\tau_i}(0|x)$  es la función de densidad de  $y$  condicional en  $x' \beta(\tau_s)$ .

Dedicado: A mi Familia y Amigos que siempre me apoyan

Bibliografía:

- Machado, Koenker y Fitzenberger (2001): *Economic Applications of Quantile Regression*, Springer-Verlag, texto de aplicaciones.
- Koenker, R. (2005): *Quantile Regression*, Cambridge University Press

*“No perdáis vuestro tiempo ni en llorar el pasado ni en llorar el porvenir. Vivid vuestras horas, vuestros minutos. Las alegrías son como flores que la lluvia mancha y el viento deshoja.”*

*Miguel de Cervantes Saavedra*