

Regresión Lineal

Rodrigo Negrete Pérez

February 2, 2022

1 Regresión a la Media

2 GDP

3 MCO

Section 1

Regresión a la Media

Chicharitos

- En 1889 Galton popularizó la “regresión a la media”
- Las vainas de chícharos:
 - Las extremadamente largas tenían hijas largas, pero no tanto como sus padres
 - Las extremadamente pequeñas tenían hijas pequeñas, pero no tanto como su padres
 - Los hijos revertían, regresaban a un promedio
- Queremos una línea que, con los datos, prediga la longitud promedio de los hijos, dada la de los padres

FEC (CEF)

- El objeto matemático es la *Función de Esperanza Condicional*
- CEF en inglés
- Nos da el promedio de Y dado un conjunto de valores de X (condicionando por X)
- Puede tener cualquier forma, pero podemos estimarla usando una línea

Section 2

GDP

¿Qué queremos?

- Nos interesa la relación entre una variable X y una variable Y
 - Un tratamiento y un outcome
- En general, no sabemos “la verdad”, las relaciones auténticas
- El objetivo es acercarnos a “la verdad” a través de los datos
- Con datos simulados sí que sabemos la verdad

Proceso de Generación de Datos (GDP)

- n observaciones con dos variables: (x_1, x_2, \dots, x_n) y (y_1, y_2, \dots, y_n)
- GDP es como un caja negra: entra una x , pasa algo, y se convierte en una y
- Definamos un GDP

$$y_i = \alpha + \beta x_i + \epsilon_i$$

- reconocemos que hay procesos estocásticos (ϵ)

Construyamos un df. Supongamos:

- $N = 10,000$
- $\alpha = 5$
- $\beta = 4$
- $x \sim N(20, 15)$
- $\epsilon \sim N(5, 2)$

```
set.seed(2022)
n<-10000
alpha=5
beta= 4
x<-rnorm(n,50,15)
e<-rnorm(n,5,2)
y<-alpha+ beta*x + e
df<-data.frame(x,y)
```

	x	y
1	63.50	266.21
2	32.40	140.07
3	36.54	155.11
4	28.33	122.40
5	45.03	191.20
6	6.49	36.62

- En nuestro caso, el GDP es la “verdad”
- Entonces, ¿cómo podemos inferir la verdad a partir de la muestra?

Section 3

MCO

Mínimos Cuadrados Ordinarios (MCO)

- Se publicó el procedimiento en 1805 por Legendre
- Gauss ya lo había empleado
 - Le pareció tan trivial que lo publicó hasta 1809
 - Desarrolló algunos teoremas para el procedimiento
- OLS en inglés

- Ya con nuestros datos, queremos la línea que encaje mejor