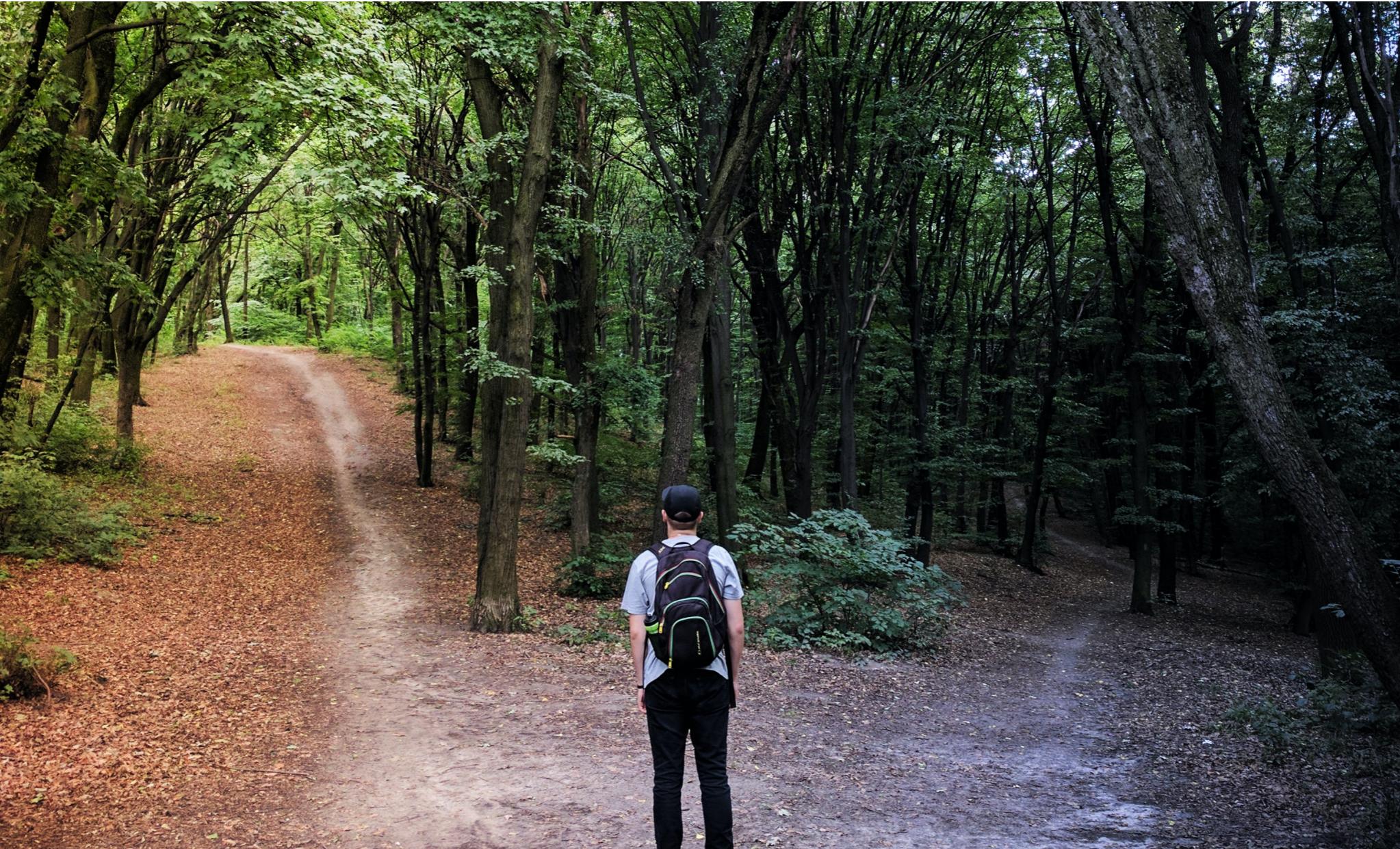


Modelos Lineales

Maestría en Estadística Aplicada

- ¿Cuál es la relación causal que nos interesa?
- ¿Cómo se vería un experimento ideal?
- ¿Cuál es nuestra estrategia de identificación?
- ¿Cuál es nuestro modo de inferencia estadística?



Experimentos Aleatorios

El modelo ideal para caminos divergentes

Health and demographic characteristics of insured and uninsured
couples in the NHIS

	Husbands			Wives		
	Some HI	No HI	Difference	Some HI	No HI	Difference
	(1)	(2)	(3)	(4)	(5)	(6)
A. Health						
Health index	4.01 [.93]	3.70 [1.01]	.31 (.03)	4.02 [.92]	3.62 [1.01]	.39 (.04)
B. Characteristics						
Nonwhite	.16	.17	-.01 (.01)	.15	.17	-.02 (.01)
Age	43.98	41.26	2.71 (.29)	42.24	39.62	2.62 (.30)
Education	14.31	11.56	2.74 (.10)	14.44	11.80	2.64 (.11)
Family size	3.50	3.98	-.47 (.05)	3.49	3.93	-.43 (.05)
Employed	.92	.85	.07 (.01)	.77	.56	.21 (.02)

TABLE 1.1
 Health and demographic characteristics of insured and uninsured
 couples in the NHIS

	Husbands			Wives		
	Some HI	No HI	Difference	Some HI	No HI	Difference
	(1)	(2)	(3)	(4)	(5)	(6)
A. Health						
Health index	4.01 [.93]	3.70 [1.01]	.31 (.03)	4.02 [.92]	3.62 [1.01]	.39 (.04)
B. Characteristics						
Nonwhite	.16	.17	−.01 (.01)	.15	.17	−.02 (.01)
Age	43.98	41.26	2.71 (.29)	42.24	39.62	2.62 (.30)
Education	14.31	11.56	2.74 (.10)	14.44	11.80	2.64 (.11)
Family size	3.50	3.98	−.47 (.05)	3.49	3.93	−.43 (.05)
Employed	.92	.85	.07 (.01)	.77	.56	.21 (.02)
Family income	106,467	45,656	60,810 (1,355)	106,212	46,385	59,828 (1,406)
Sample size	8,114	1,281		8,264	1,131	

Notes: This table reports average characteristics for insured and uninsured married couples in the 2009 National Health Interview Survey (NHIS). Columns (1), (2), (4), and (5) show average characteristics of the group of individuals specified by the column heading. Columns (3) and (6) report the difference between the average characteristic for individuals with and without health insurance (HI). Standard deviations are in brackets; standard errors are reported in parentheses.

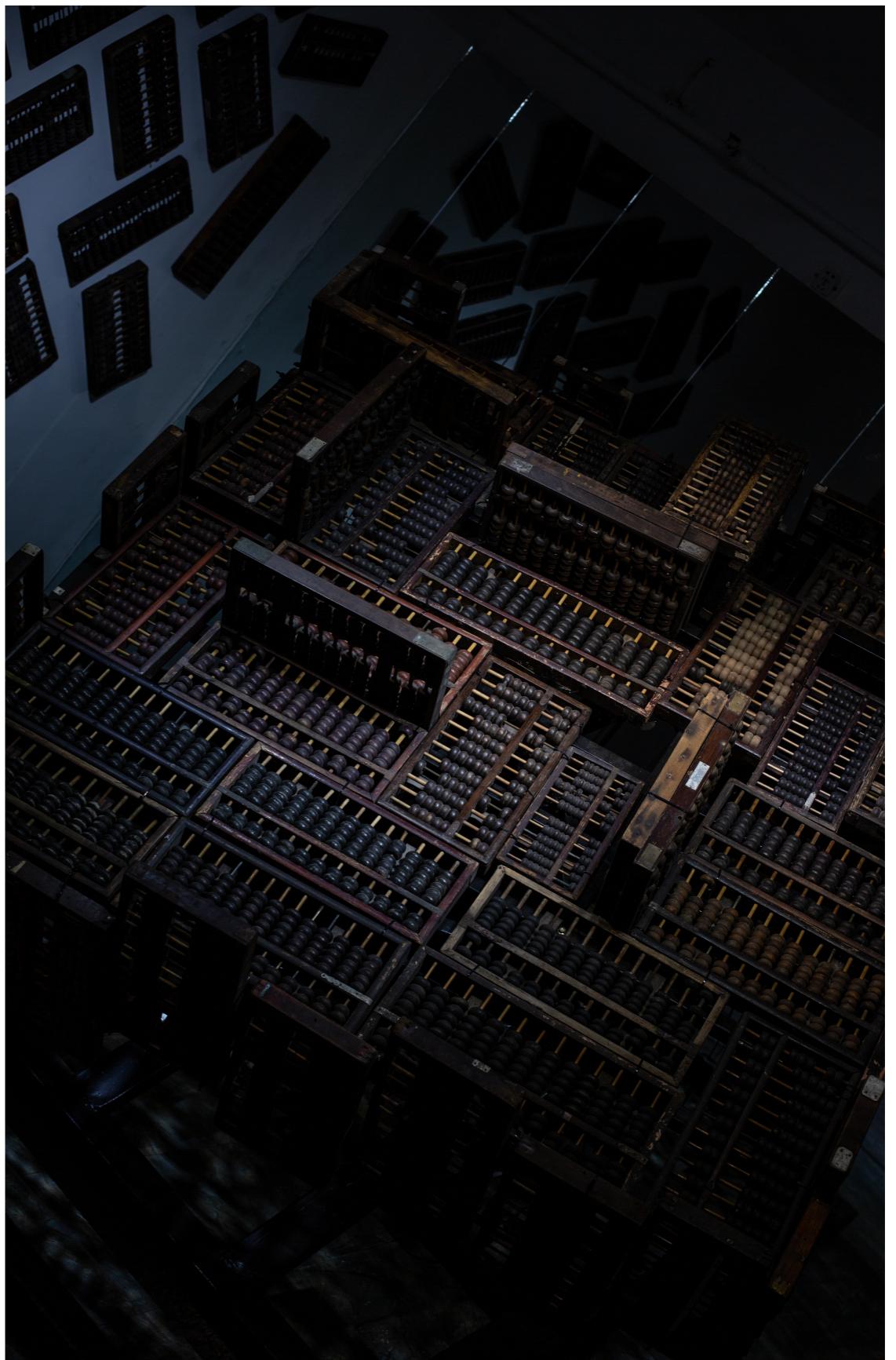
$$Y_{1,Angelina}-Y_{0,Angelina}=1$$

$$Y_{1,Daniel}-Y_{0,Daniel}=0$$

$$Y_{Angelina} - Y_{Daniel} = -1$$

$$\begin{aligned} Y_{Angelina}-Y_{Daniel}&=Y_{1,Angelina}-Y_{0,Daniel}\\ &=Y_{1,A}-Y_{0,A}+Y_{0,A}-Y_{0,D} \end{aligned}$$

Repaso de conceptos



$$\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i$$

$$\sum_{i=1}^n (x_i - \bar{x}) = 0$$



$$\sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n x_i^2 - n(\bar{x})^2$$



Valor Esperado

$$E(X) = x_1 f(x_1) + x_2 f(x_2) + \cdots + x_k f(x_k)$$

$$= \sum_{j=1}^k x_j f(x_j)$$

Propiedades del valor esperado

$$E(aX + b) = aE(X) + b$$

$$E(a_1X_1 + \dots + a_nX_n) = a_1E(X_1) + \dots + a_nE(X_n)$$

$$E\left(\sum_{i=1}^n a_iX_i\right) = \sum_{i=1}^n a_iE(X_i)$$

Más propiedades

$$E(c) = c$$

$$E(W + H) = E(W) + E(H)$$

$$E(W - E(W)) = 0$$

Varianza

$$V(X) = E[(X - E[X])^2]$$

$$V(W) = E(W^2) - E(W)^2$$

Varianza muestral

$$\hat{S}^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{(n - 1)}$$

Propiedades

$$V(aX + b) = a^2 V(X)$$

$$V(c) = 0$$

Varianza de la suma de dos v.a.

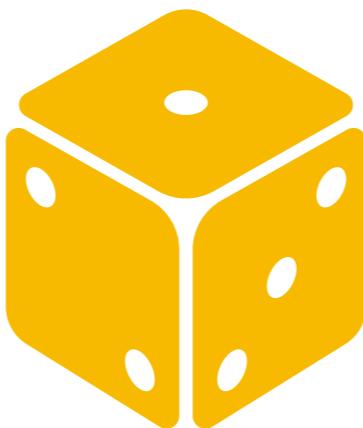
$$V(X + Y) = V(X) + V(Y) + 2 \left(E(XY) - E(X)E(Y) \right)$$

¿Demostración?



Ejercicio: Obtén el valor esperado y la varianza de un dado justo.

(Y obtén una fórmula general para un dado con n lados)



Covarianza

$$C(X, Y) = E(XY) - E(X)E(Y)$$

Prove it!!

Covarianza

$$V(X) = \text{Cov}(X, X)$$

$$C(a_1 + b_1 X, a_2 + b_2 Y) = b_1 b_2 C(X, Y)$$

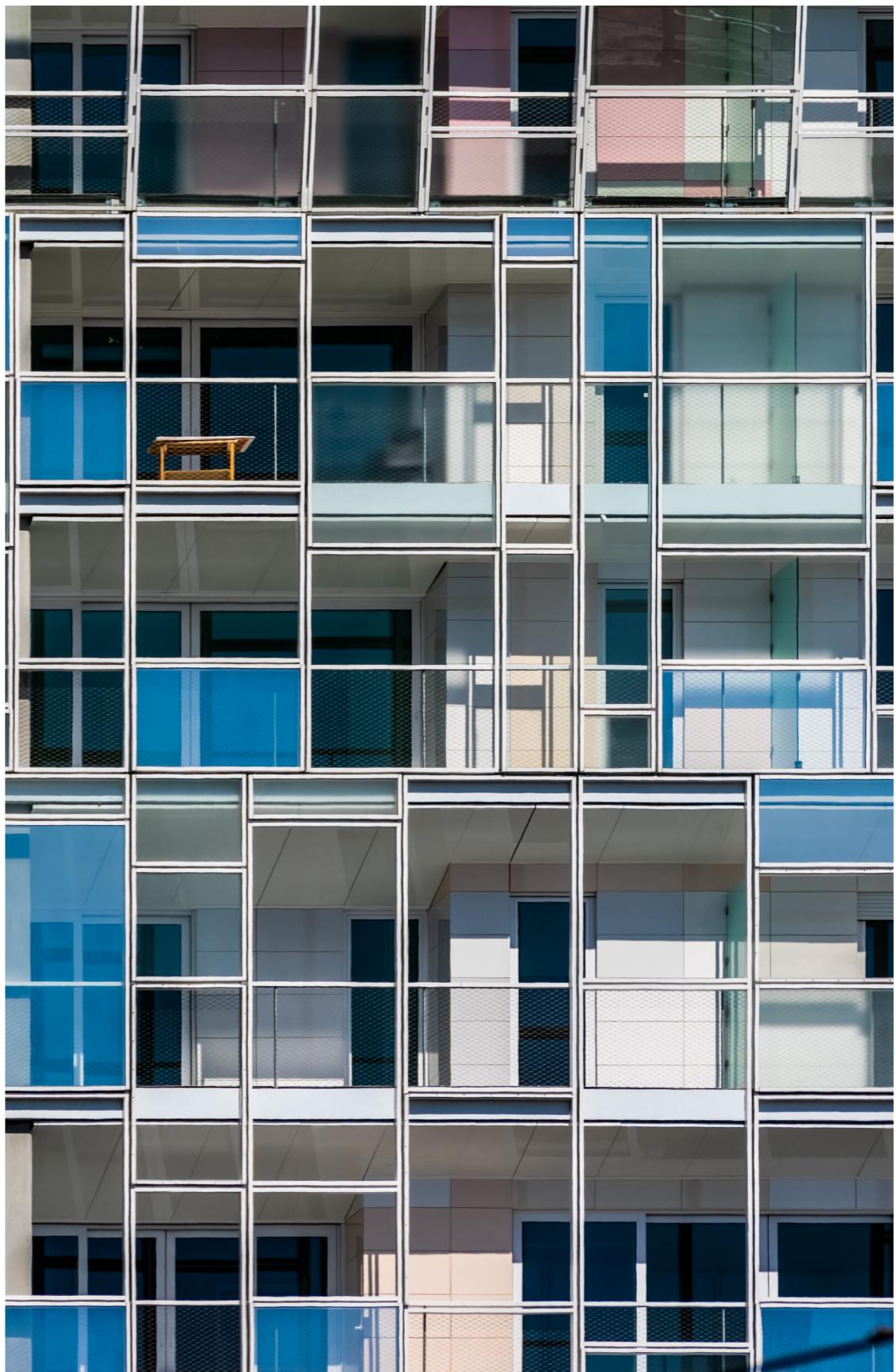
Ejercicio: Demostrar

Correlación

$$\rho_{X,Y} = \frac{C(X, Y)}{\sigma_X \sigma_Y}$$

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

Mínimos Cuadrados Ordinarios



Modelo lineal de población

$$y = \beta_0 + \beta_1 x + u$$

$$E(u) = 0$$

Mínimos cuadrados ordinarios

Sean $\{(x_i, y_i) : u = 1, 2, \dots, n\}$ muestras aleatorias
de tamaño n de la población

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

Independencia media

$$E(u | x) = E(u)$$

Para todos los valores de x

$$E(u | x) = 0$$

Para todo valor de x

No tenemos acceso a la población, pero si a sus contrapartes en muestra

$$\frac{1}{n} \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$$

$$\frac{1}{n} \sum_{i=1}^n x_i [y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i] = 0$$

Estimando los parámetros

$$\frac{1}{n} \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)$$

$$= \frac{1}{n} \sum_{i=1}^n y_i - \frac{1}{n} \sum_{i=1}^n \hat{\beta}_0 - \frac{1}{n} \sum_{i=1}^n \hat{\beta}_1 x_i$$

$$= \bar{y} - \hat{\beta}_0 - \hat{\beta}_1 \bar{x} = 0$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Estimando $\hat{\beta}_1$

$$E(u \mid x) = 0$$

$$(x[y - \beta_0 - \beta_1 x] = 0$$

En la muestra

$$\frac{1}{n} \sum_{i=1}^n x_i(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$$

$$\hat{\beta}_1 = \frac{\mathbf{Cov}(x_i, y_i)}{\sigma^2(x_i)}$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Ejercicio: Demostrar

$$E(y \mid x) = \beta_0 + \beta_1 x$$

Suma de los errores al cuadrado

$$\sum_{i=1}^n \hat{u}_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

$$= \sum_{i=1}^n \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i \right)^2$$

Buenosidad de Ajuste

$$SST = \sum_{i=1}^n (y_i - \bar{y})^2$$

$$SSE = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$$

$$SSR = \sum_{i=1}^n \hat{u}_i^2$$

Propiedades estadísticas de los MCO



Valor esperado de MCO

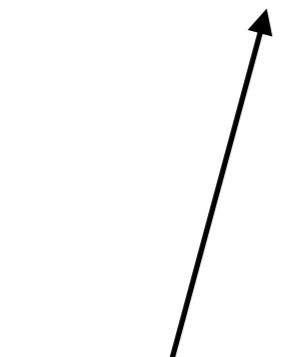
$$E(\hat{\beta}) = \beta$$

Supuestos de un MCO *insesgado*

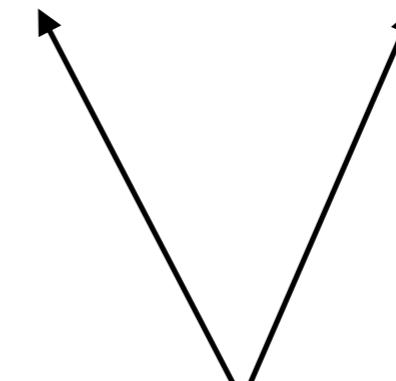
- Lineal en los parámetros
- Muestreo aleatorio
- Variación muestral en la variable explicatoria (la varianza muestral de x no es cero)
- Media condicional cero $E(u | x) = 0$

Linealidad en los parámetros

$$y = \beta_0 + \beta_1 x + u$$



Aleatorio



Aleatorio

Muestreo aleatorio

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

u_i es un error no observado para i , no el residuo

Variación muestra en la variable explitativa

$$\hat{\sigma}^2(x) > 0$$

Media condicional cero

$$E(u | x) = E(u) = 0$$

$$\implies E(\hat{\beta}_1) = \beta_1$$

Instrucciones para demostrar
que $\hat{\beta}_1$ es insesgado

Ley de las esperanzas iteradas



Conditional Expectation Function (CEF)

Es la media de algún resultado y con alguna covariable x fija

$$E(y_i | x_i)$$

Law of iterated expectations

$$E(y_i) = E\{E(y_i | x_i)\}$$

Demostración

$$E\{E(y|x)\} = \int E(y|x=u)g_x(u)du$$

$$= \int \left[\int t f_{y|x}(t|x=u) dt \right] g_x(u) du$$

$$\int \int t f_{y|x}(t \mid x = u) g_x(u) du dt$$

$$\int t \left[\int f_{y|x}(t \mid x = u \mid g_x(u) du \right] dt$$

$$= \int t [f_{x,y} du] dt$$

$$= \int t g_y(t) dt$$

$$= E(y)$$