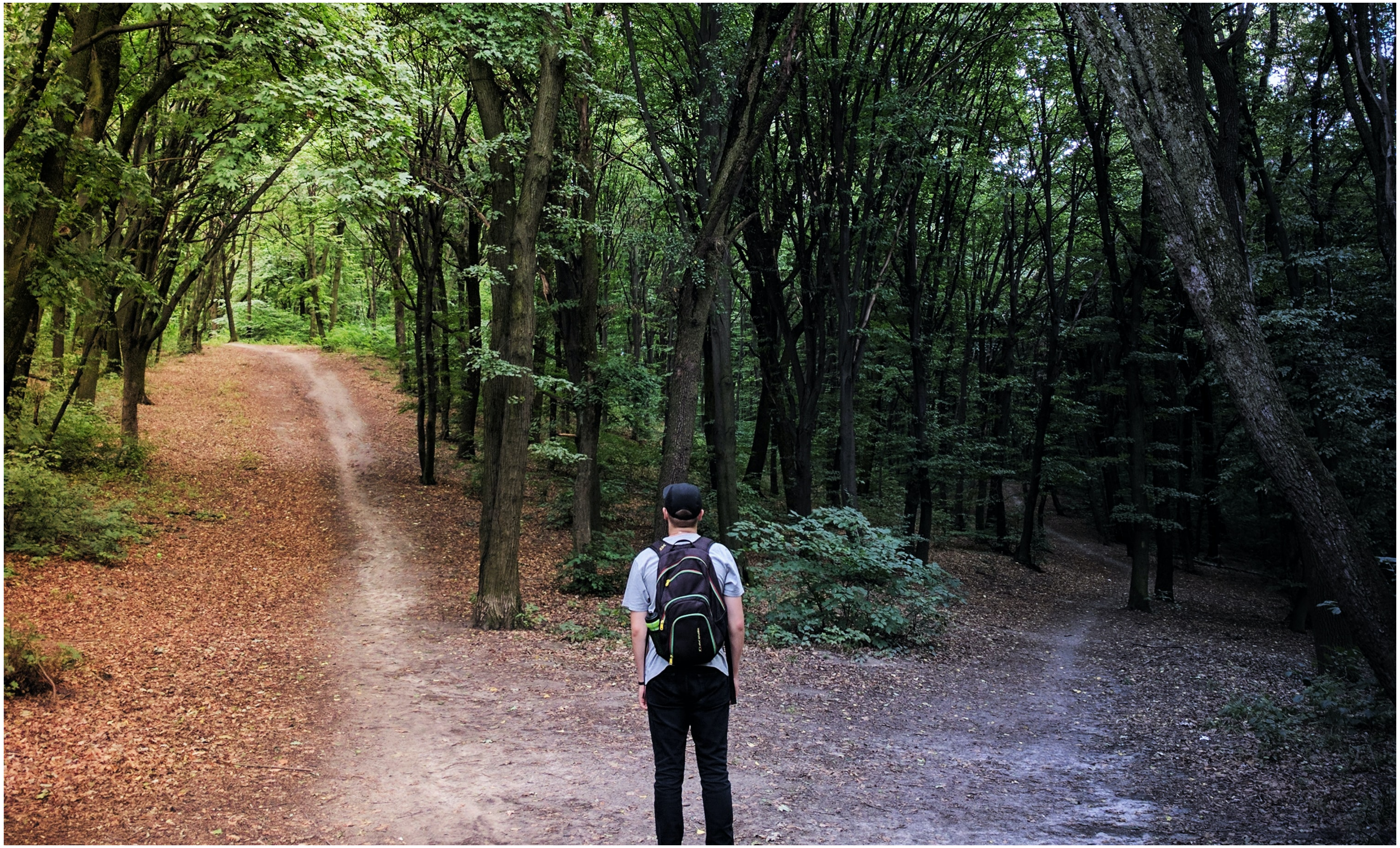
Modelos Lineales

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

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

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

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

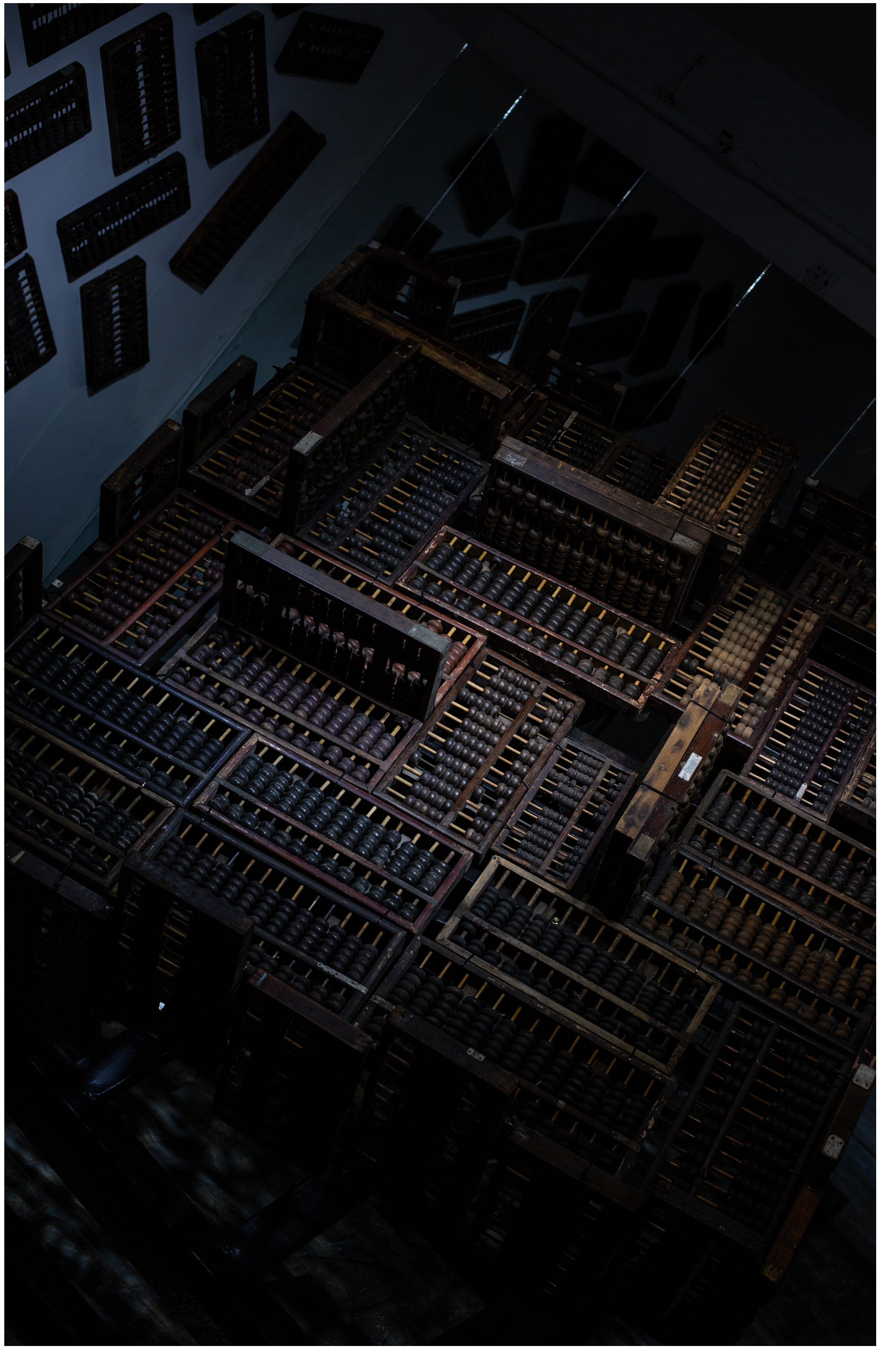
$$Y_{1,Daniel} - Y_{0,Daniel} = 0$$

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

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

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

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

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

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

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

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

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

$$V(c) = 0$$

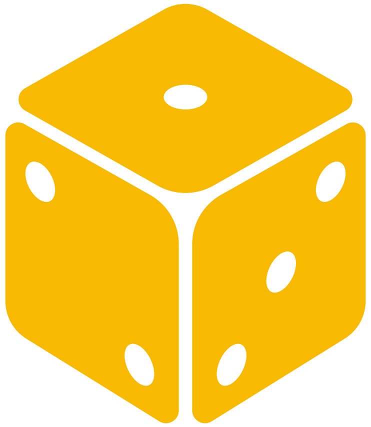
Varianza de la suma de dos v.a.

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

¿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

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

Covarianza

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

$$C(a_1 + b_1X, a_2 + b_2Y) = b_1b_2C(X, Y)$$

Ejercicio: Demostrar

Correlación

Sean

$$W = \frac{X - E(X)}{\sqrt{V(X)}} \qquad Z = \frac{Y - E(Y)}{\sqrt{V(Y)}}$$

$$\textbf{Corr}(W, Z) = \frac{C(X, Y)}{\sqrt{V(X)V(Y)}}$$

Modelo lineal de población

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

$$E(u) = 0$$

Independencia media

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

Para todos los valores de x

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

Para todo valor de x

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

Mínimos cuadrados ordinarios

Sean $\{(x_i, y_i : u = 1,2,\ldots,n\}$ muestras aleatorias

de tamaño n de la población

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

$$E(u) = 0$$

$\therefore$

$$E(y - \beta_0 - \beta_1 x) = 0$$

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

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

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

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

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

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

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

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

$$\hat{\beta}_1 = \frac{\textbf{Cov}(x_i, y_i)}{\textbf{Cov}(x_i)}$$

Ejercicio: Demostrar