



Métodos Cuantitativos para Asuntos Públicos I

EGOB 2101

Andrés Ham
a.ham@uniandes.edu.co

2019-1

13 de febrero del 2019

Agenda de hoy

- 1 Correlación: Cosas que se mueven juntas
- 2 Calculando la correlación
- 3 Interpretando la correlación
- 4 El mensaje del día

Previously...

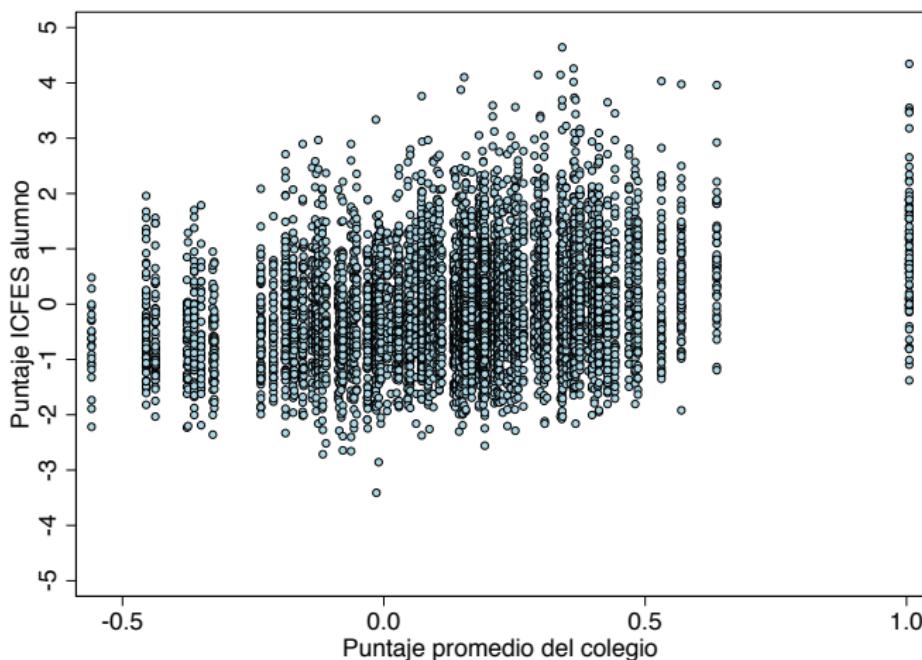
- ▶ La estadística descriptiva nos permite aprender mucho sobre un todo (la población) utilizando una parte (la muestra).
- ▶ Pero, calcular promedios y realizar tests de hipótesis sobre esas medias son sólo la punta del iceberg. *Hay mucho por aprender.*
- ▶ Gran parte del análisis cuantitativo consiste en buscar y evaluar patrones, es decir, cómo interactúan diferentes variables entre sí.

En nuestro ejemplo ICFES...

- ▶ ...no sería loco pensar que el puntaje individual en el ICFES está relacionado con el rendimiento general del colegio.

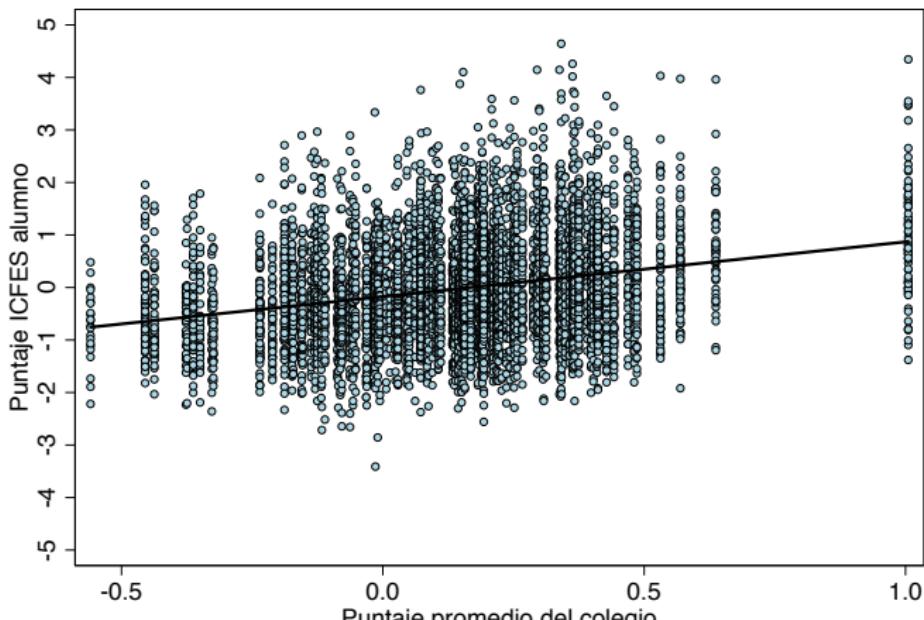
En nuestro ejemplo ICFES...

- ▶ ...no sería loco pensar que el puntaje individual en el ICFES está relacionado con el rendimiento general del colegio.



En nuestro ejemplo ICFES...

- ▶ Si ajustamos una línea a esos datos:



¿Qué es la correlación?

- ▶ Cuando dos variables se mueven juntas, decimos que están **correlacionadas**.

¿Qué es la correlación?

- ▶ Cuando dos variables se mueven juntas, decimos que están **correlacionadas**.
- ▶ Hay dos tipos de correlación:

¿Qué es la correlación?

- ▶ Cuando dos variables se mueven juntas, decimos que están **correlacionadas**.
- ▶ Hay dos tipos de correlación:
 - 1 **Positiva**: valores grandes de la variable 1 ocurren cuando hay valores grandes de la variable 2.

¿Qué es la correlación?

- ▶ Cuando dos variables se mueven juntas, decimos que están **correlacionadas**.
- ▶ Hay dos tipos de correlación:
 - 1 *Positiva*: valores grandes de la variable 1 ocurren cuando hay valores grandes de la variable 2.
 - 2 *Negativa*: valores grandes de la variable 1 ocurren cuando hay valores pequeños de la variable 2.

Una historia muy Latinoamericana

Honduran President Declared Winner, but O.A.S. Calls for New Election

[Leer en español](#)

By ELISABETH MALKIN DEC. 17, 2017



¿Qué paso en Honduras?

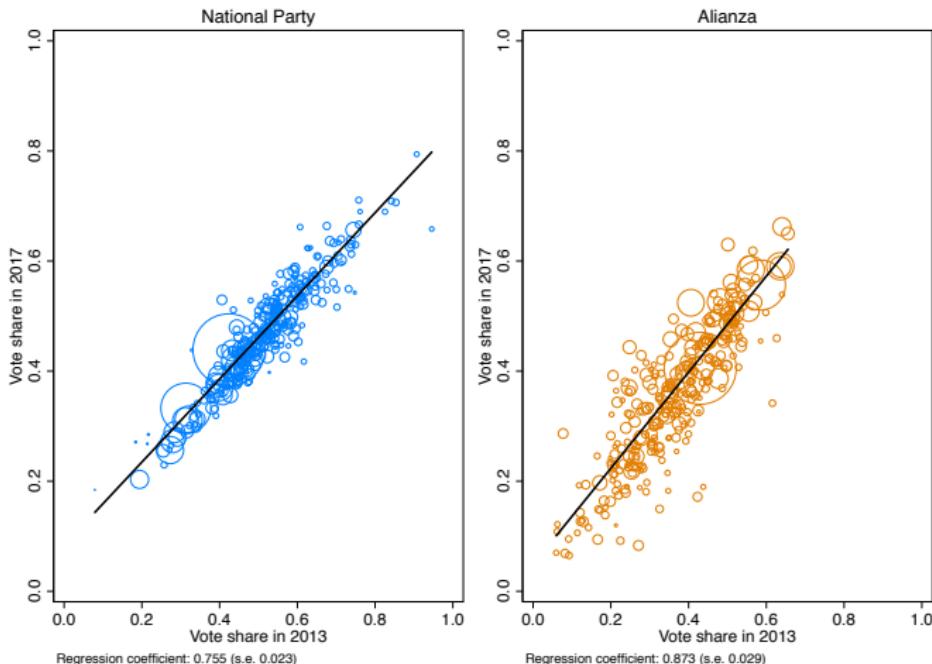
- ▶ El presidente logró legitimizar la reelección, pese a estar Constitucionalmente prohibida.
- ▶ Las elecciones fueron algo extrañas, por diversas razones.
- ▶ Por suerte, el Tribunal Supremo Electoral subió sus datos a internet, lo cuál permite analizarlos para descifrar qué paso.

Correlación positiva

- ▶ ¿Qué tan fiel es la gente a su partido?

Correlación positiva

- ▶ ¿Qué tan fiel es la gente a su partido?



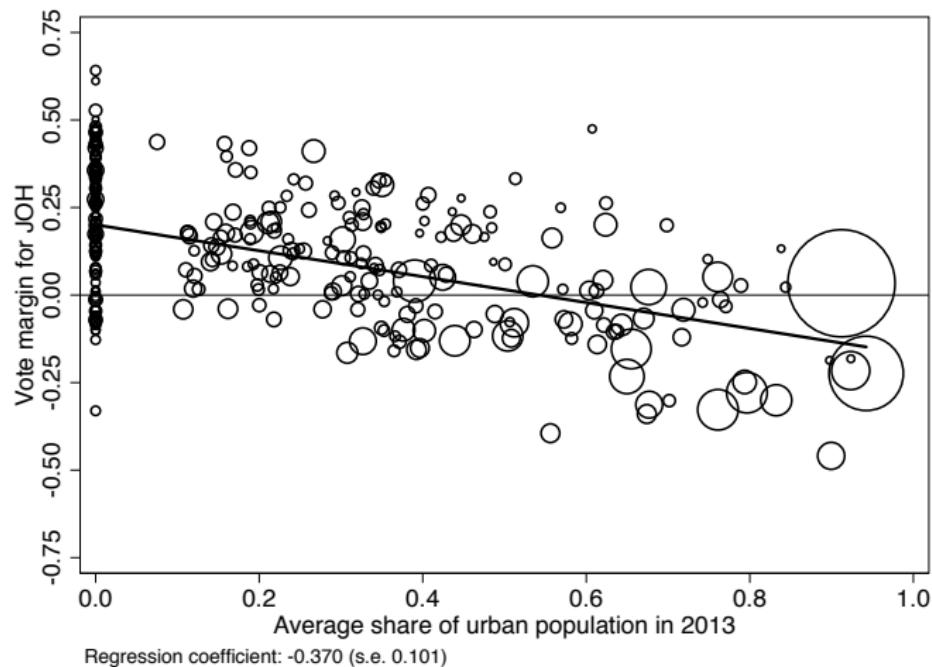
Fuente: Ham (2019).

Correlación negativa

- ▶ El presidente dice que ganó en zonas rurales. ¿Será cierto?

Correlación negativa

- ▶ El presidente dice que ganó en zonas rurales. ¿Será cierto?



Fuente: Ham (2019).

Los gráficos son bonitos, pero...

- ▶ No podemos presentar todos los gráficos, aunque se vean lindos.

Los gráficos son bonitos, pero...

- ▶ No podemos presentar todos los gráficos, aunque se vean lindos.
- ▶ Hay medidas que nos permiten *resumir* las relaciones estadísticas entre variables.

Los gráficos son bonitos, pero...

- ▶ No podemos presentar todos los gráficos, aunque se vean lindos.
- ▶ Hay medidas que nos permiten *resumir* las relaciones estadísticas entre variables.
- ▶ Para determinar si dos variables están asociadas (y su dirección), podemos calcular el **coeficiente de correlación**.

La bestia

- ▶ El coeficiente de correlación se calcula con la siguiente fórmula:

La bestia

- ▶ El coeficiente de correlación se calcula con la siguiente fórmula:

$$\rho = \frac{1}{n} \sum_{i=1}^n \frac{(x_i - \bar{x})}{\sigma_x} \frac{(y_i - \bar{y})}{\sigma_y}$$

La bestia

- ▶ El coeficiente de correlación se calcula con la siguiente fórmula:

$$\rho = \frac{1}{n} \sum_{i=1}^n \frac{(x_i - \bar{x})}{\sigma_x} \frac{(y_i - \bar{y})}{\sigma_y}$$

- ▶ n es el número de observaciones.
- ▶ \bar{x} es el promedio de la variable 1.
- ▶ \bar{y} es el promedio de la variable 2.
- ▶ σ_x es el desvío estándar de la variable 1.
- ▶ σ_y es el desvío estándar de la variable 2.

La bestia

- ▶ El coeficiente de correlación se calcula con la siguiente fórmula:

$$\rho = \frac{1}{n} \sum_{i=1}^n \frac{(x_i - \bar{x})}{\sigma_x} \frac{(y_i - \bar{y})}{\sigma_y}$$

- ▶ n es el número de observaciones.
- ▶ \bar{x} es el promedio de la variable 1.
- ▶ \bar{y} es el promedio de la variable 2.
- ▶ σ_x es el desvío estándar de la variable 1.
- ▶ σ_y es el desvío estándar de la variable 2.
- ▶ No es necesario que se la aprendan de memoria, pero sí que entiendan cómo funciona.

¿Qué nos provee este cálculo?

- ▶ Un número entre -1 y 1.

¿Qué nos provee este cálculo?

- ▶ Un número entre -1 y 1. ¿Por qué esos valores?

¿Qué nos provee este cálculo?

- ▶ Un número entre -1 y 1. ¿Por qué esos valores?
 - ▶ Si es negativo, las variables se mueven en direcciones opuestas.
 - ▶ Si es positivo, las variables se mueven en la misma dirección.
 - ▶ Si es = 0, las dos variables no están relacionadas.

¿Qué nos provee este cálculo?

- ▶ Un número entre -1 y 1. ¿Por qué esos valores?
 - ▶ Si es negativo, las variables se mueven en direcciones opuestas.
 - ▶ Si es positivo, las variables se mueven en la misma dirección.
 - ▶ Si es = 0, las dos variables no están relacionadas.
- ▶ En qué unidades está expresado este número?

¿Qué nos provee este cálculo?

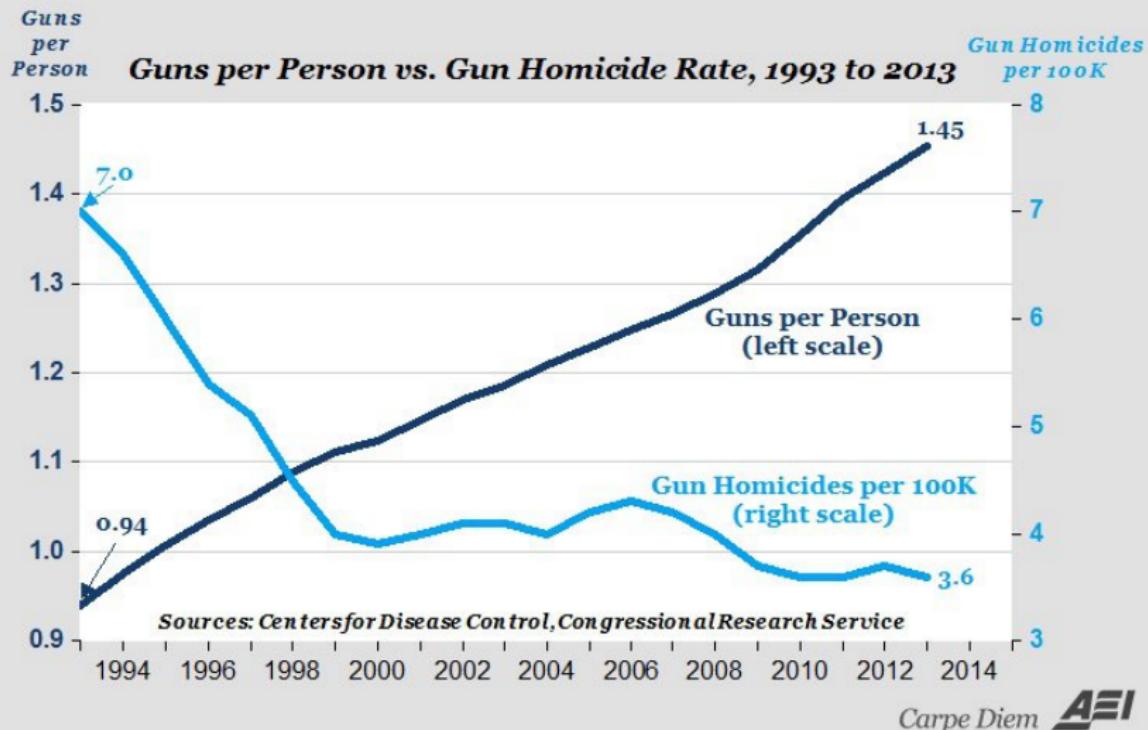
- ▶ Un número entre -1 y 1. ¿Por qué esos valores?
 - ▶ Si es negativo, las variables se mueven en direcciones opuestas.
 - ▶ Si es positivo, las variables se mueven en la misma dirección.
 - ▶ Si es = 0, las dos variables no están relacionadas.
- ▶ En qué unidades está expresado este número? ¿Por qué importa?

Algunas correlaciones de los datos del ICFES

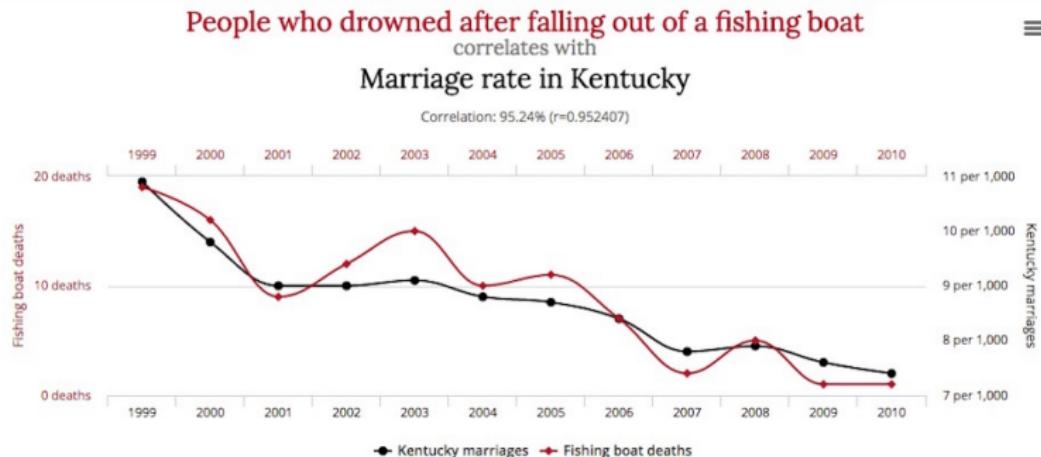
	Coeficiente de correlación
Hombres	0.172
Edad	-0.161
Educación de los padres	0.207
Ingreso familiar	0.162
Jornada matutina	0.047
Puntaje ICFES del colegio	0.259
Número de alumnos en el colegio	0.021

Calculos realizados utilizando datos de Bonilla, Bottan y Ham (2017).

Muchas cosas están correlacionadas



Muchas cosas están correlacionadas

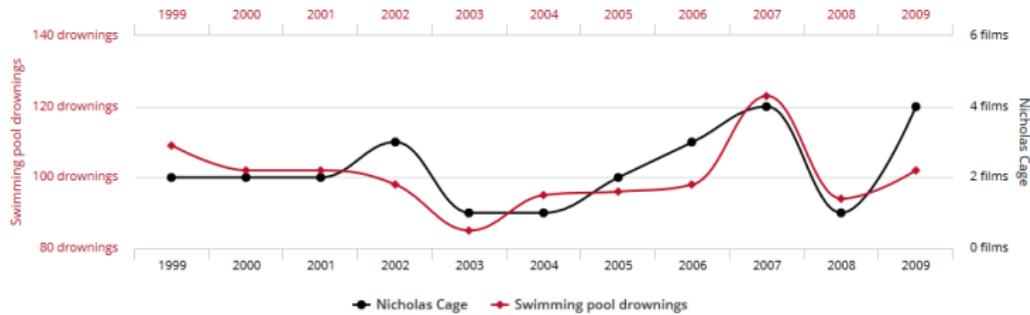


Data sources: Centers for Disease Control & Prevention and National Vital Statistics Reports

Muchas cosas están correlacionadas

Number of people who drowned by falling into a pool
correlates with
Films Nicolas Cage appeared in

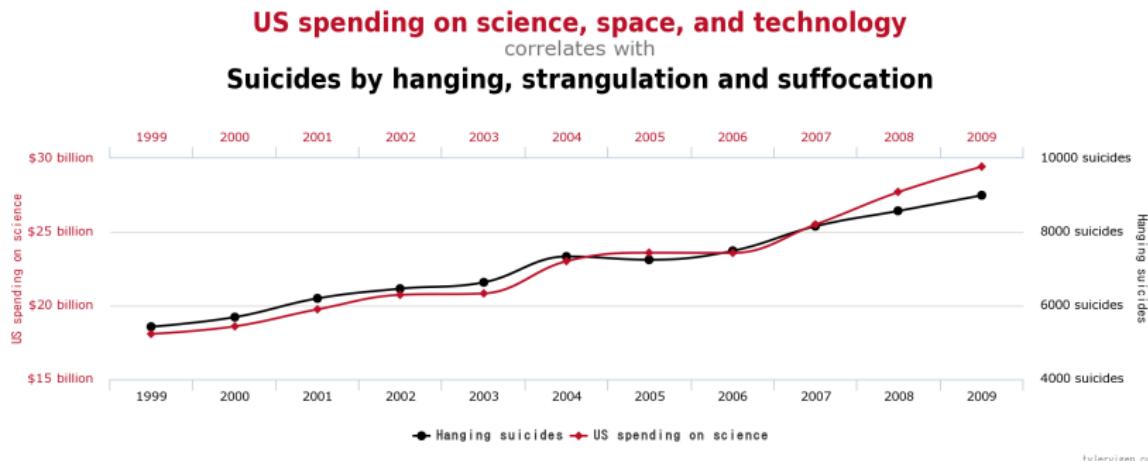
Correlation: 66.6% ($r=0.6666004$, $p>0.05$)



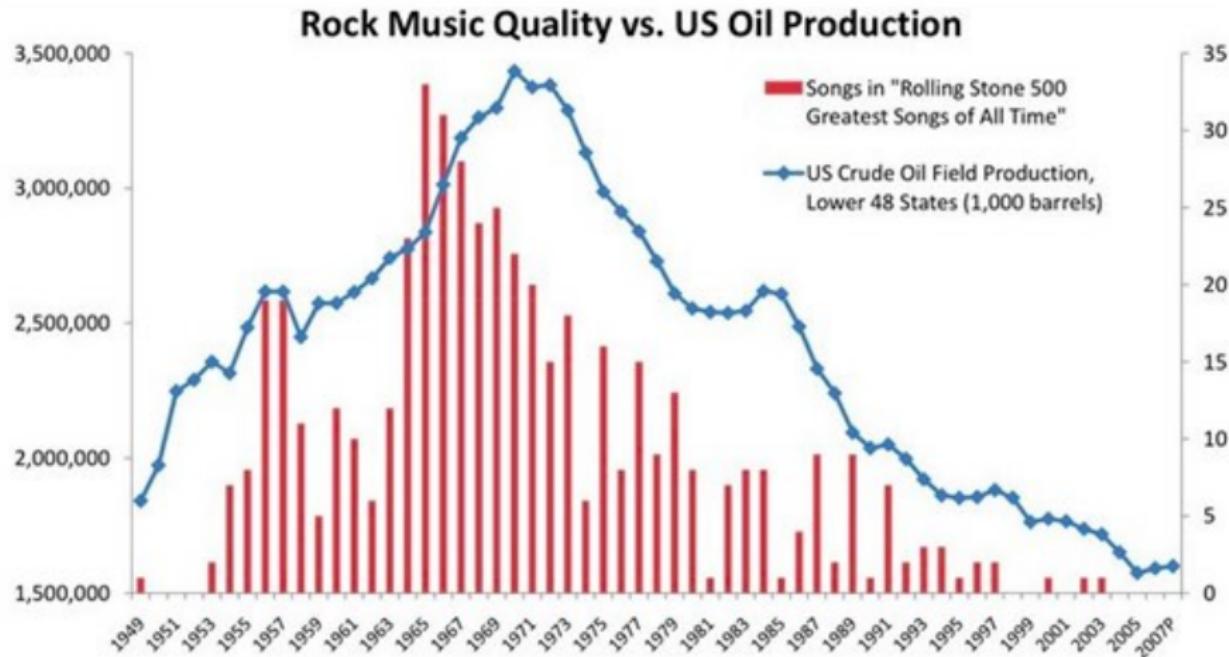
Data sources: Centers for Disease Control & Prevention and Internet Movie Database

tylervigen.com

Muchas cosas están correlacionadas



Muchas cosas están correlacionadas



Muchas cosas están correlacionadas

- ▶ Pero no todas las correlaciones proveen información útil.

Muchas cosas están correlacionadas

- ▶ Pero no todas las correlaciones proveen información útil.
- ▶ Solamente porque dos cosas se muevan juntas, no significa que una implica a la otra y viceversa.

Muchas cosas están correlacionadas

- ▶ Pero no todas las correlaciones proveen información útil.
- ▶ Solamente porque dos cosas se muevan juntas, no significa que una implica a la otra y viceversa.
- ▶ La correlación puede darse por muchos factores: causalidad, explicación parcial, y suerte.

Correlación y causalidad

1 Correlación = causa: una variable causa a la otra.

Correlación y causalidad

- 1 Correlación = causa:** una variable causa a la otra.
- 2 Correlación parcial:** una variable es una de las muchas causas de la otra variable.

Correlación y causalidad

- 1 Correlación = causa:** una variable causa a la otra.
- 2 Correlación parcial:** una variable es una de las muchas causas de la otra variable.
- 3 Correlación espuria:** las variables están asociadas por pura suerte.

Correlación y causalidad

- ▶ Esto significa que si bien encontramos relaciones en los datos, establecer causa y efecto no es una tarea simple.

Correlación y causalidad

- ▶ Esto significa que si bien encontramos relaciones en los datos, establecer causa y efecto no es una tarea simple.
- ▶ De hecho, la mayoría de lo que queda del curso va a tratar de encontrar situaciones dónde sí podemos establecer links causales.

Correlación y causalidad

- ▶ Esto significa que si bien encontramos relaciones en los datos, establecer causa y efecto no es una tarea simple.
- ▶ De hecho, la mayoría de lo que queda del curso va a tratar de encontrar situaciones dónde sí podemos establecer links causales.
- ▶ Usualmente calculamos correlaciones como un segundo paso, porque buscamos historias dentro de los datos. Su misión es diferenciar ficción de realidad en la práctica.

El mensaje del día

- ▶ Dos variables están correlacionadas cuando se mueven juntas.
- ▶ Estos movimientos conjuntos se pueden describir gráficamente o resumir calculando un coeficiente de correlación.
- ▶ La correlación tiene un espectro de poder informativo, que abarca desde relaciones causales hasta espuriedad.
- ▶ Si bien es útil como herramienta, es importante reconocer qué puede y qué no puede decir.

El mensaje del día, versión Twitter



Walter Sosa Escudero

@wsosaescudero

Following



- 1) Falacia de la correlacion: crer que correlacion implica causalidad.
- 2) Hiperfalacia de la correlacion: creer que la falacia de la correlacion implica que las correlaciones no sirven para nada.
- 3) Falacia de la causalidad: creer que causalidad implica correlacion.

Translate from Spanish

2:24 PM - 29 Dec 2017

En el próximo capítulo...

- ▶ ¿Cómo mejoramos un análisis de correlación simple? Regresión.
- ▶ Lecturas:
 - ▶ Capítulo 11 de Wheelan: “*Naked Statistics*”, páginas 185-198.
 - ▶ Capítulo 2 de Angrist y Pischke: “*Mastering Metrics*”, páginas 47-68.