

Big data: desafíos y oportunidades para la economía y las ciencias sociales

Walter Sosa Escudero

Universidad de San Andrés y CONICET

wsosa@udesa.edu.ar • [@wsosaescudero](https://twitter.com/wsosaescudero) • waltersosa.weebly.com

SCIENCE : DISCOVERIES 

The End of Theory: The Data Deluge Makes the Scientific Method Obsolete

By Chris Anderson  06.23.08



2014 Significance Lecture

The Big Data Trap

Tim Hardford, Economist, journalist and broadcaster
Chair: Brian Tarran, Editor, *Significance*



DATA | EVIDENCE | DECISIONS



1:36 / 58:08



RSS 2014 Significance Lecture - The Big Data trap



The Seven Pillars of Statistical Wisdom

STEPHEN M. STIGLER



found after an accident that he could remember absolutely everything. He could reconstruct every day in the smallest detail, and he could even later reconstruct the reconstruction, but he was incapable of understanding. Borges wrote, “To think is to forget details, generalize, make abstractions. In the teeming world of Funes there were only details.”² Aggregation can yield great gains above the individual components. Funes was big data without Statistics.

When was the arithmetic mean first used to summarize a data set, and when was this practice widely adopted? These are two very different

Clarín.com > iEco > Economía > 06/04/14

Big data: ¿Otra vez arroz?

DEBATE

2 opiná



142 shares



11



131



8



1



+

**Walter Sosa
Escudero
Profesor
Asociado, Udesa**

Creo conservar, en algún recóndito lugar de mi casa, mi paleta de paddle de cuando en los noventa pensaba que el juego del presente se transformaría en el deporte del futuro. También disfruto de los vinilos de mi adolescencia que escucho casi a diario. Y por alguna razón exótica guardo celosamente una caja de diskettes de mis comienzos con la computación personal, allá en los ochenta.

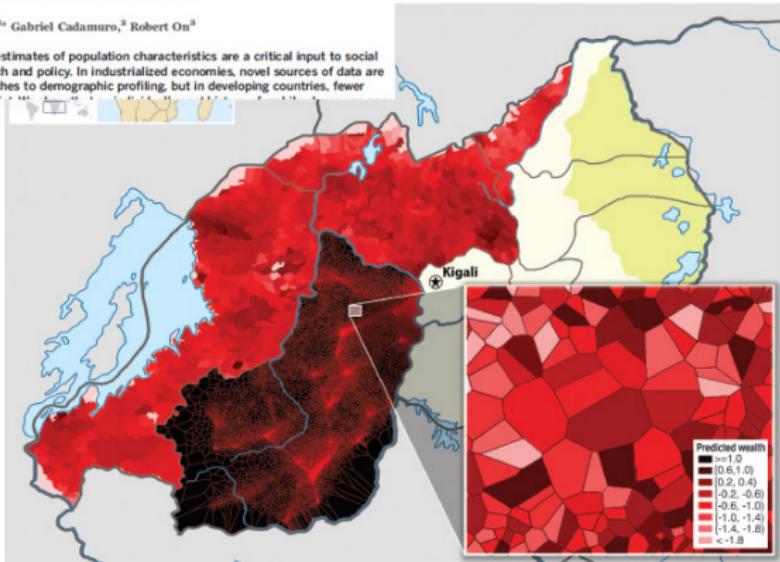
Pobreza en Rwanda (predecir)

ECONOMICS

Predicting poverty and wealth from mobile phone metadata

Joshua Blumenstock,^{1*} Gabriel Cadamuro,² Robert On³

Accurate and timely estimates of population characteristics are a critical input to social and economic research and policy. In industrialized economies, novel sources of data are enabling new approaches to demographic profiling, but in developing countries, fewer



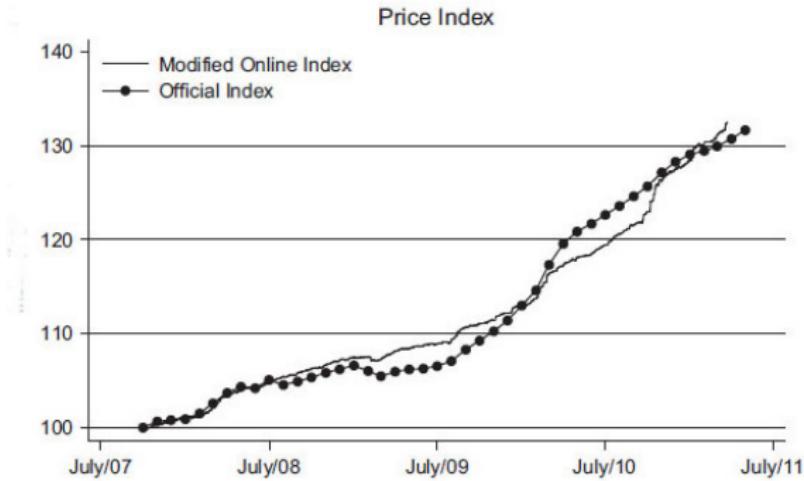
Precios en Argentina (medir)



Online and official price indexes: Measuring Argentina's inflation

Alberto Cavallo *

Massachusetts Institute of Technology, Sloan School of Management, 27 Massachusetts Ave 32G-312, Cambridge, MA 02139, USA



Sales Taxes and Internet Commerce

Liran Einav

Dan Knoepfle

Jonathan Levin

Neel Sundareshan

AMERICAN ECONOMIC REVIEW
VOL. 104, NO. 1, JANUARY 2014
(pp. 1-26)

Small data (estadistica clasica)

- Extraer lo maximo de **pocos** datos
- Solucion: **estructurar** los datos (muestreo)
- Enfoque: muestreo complejo aproxima muestreo al azar (**lento** y **caro**, pero bueno).

Big data

- **Muchos** datos (Volumen)
- Muchos datos **no estructurados** (Variedad)
- Muchos datos no estructurados e **immediatos** (Velocidad)
- 'Condisionalmente baratos'.

$$Y = f(X) + u$$

Estadistica clasica:

- Interes en $f(.)$. Efecto causal
- Modelo: Teoria, experimento.
- Probabilidades (error estandar, tests)
- Bueno?: insesagdo, varianza minima, inferencia valida.

$$Y = f(X) + u$$

Machine learning:

- Interes en Y : predecir, clasificar, medir.
- Modelo: modelo?. Lo **aprendemos**.
- Prediccion puntual (no inferencia).
- Bueno?: Performance predictiva fuera de la muestra.

Ejemplo: Ridge / LASSO

$$L(\beta) = \sum (y_i - x_i\beta)^2 + \lambda \beta^2$$

- $\lambda = 0$ de vuelta a MCO.
- $\lambda \neq 0$ sesgado pero...
- ... **siempre** le puede ganar a MCO en terminos de predicción.
- Sesgo: pecado mortal.
- **ML:** sesgo puede reducir la varianza dramaticamente.
- Puede lidiar con $K > n$ (la regla mas que la excepción en big data).

Esto es ridge. LASSO reemplaza β^2 por $|\beta|$.

- Etiqueta estadistica: ex-ante. Teoria, identificacion 'limpia'(consistencia). Inferencia robusta.
- Machine learning: ex-post, iterativa. Cross validation.
- Machine learning **construye** el modelo mas que lo estima, en base a la performance predictiva **fuerza de la muestra**. Adios al R^2 (y a MCO? Mmm...).

- Dependencias (realmente tenemos big data?. Trump effect)
- Choice based sampling.
- Contrafactuales nunca se observan (podemos tener *todos* los datos?).
- Falacia de la correlacion.
- Transparencia / privacidad.
- Comunicabilidad. Caja negra (deep learning, forests, etc.)
- Consenso social/politico.

Oportunidades

- More data: Big data no es solo muchos sino **mas** datos (pocos o muchos).
- Identificar la complejidad/heterogeneidad. No linealidades. Maldicion de la dimensionalidad.
- Rapido (crucial para la politica). Google Flu Trends. Price scrapping.
- Oportunidad para diseñar experimentos.
- Complementa a las estadisticas oficiales (no reemplaza).
- Cobertura (Rwanda). Rural vs. urbano, etc..

En pleno auge, la "datanomics" ya tiene sus propias tribus

09 ABR 2017



La postura frente al uso de datos genera "grupos de economistas"; estructurales, experimentalistas, temporalistas, descriptivos y analistas son algunos de ellos

Por Walter Sosa Escudero

Por qué los economistas llegaron tarde al big data

BIG DATA

24 DIC 2016



Balance 2016. Encuestas y big data: no matemos al mensajero

ESPECIAL IDEAS

18 DIC 2016



Criticados por sus errores de pronóstico, los sondeos y modelos estadísticos tienen un rol más relevante: medir grados de incertidumbre

Por Walter Sosa Escudero

Pornoimpuestos, transparencia y bienestar: el caso de Noruega

07 AGO 2016



Al infinito y más allá: Funes, Borges y big data

12 JUN 2016



Las encuestas políticas truchas y el club de los economistas que "p-hackean"

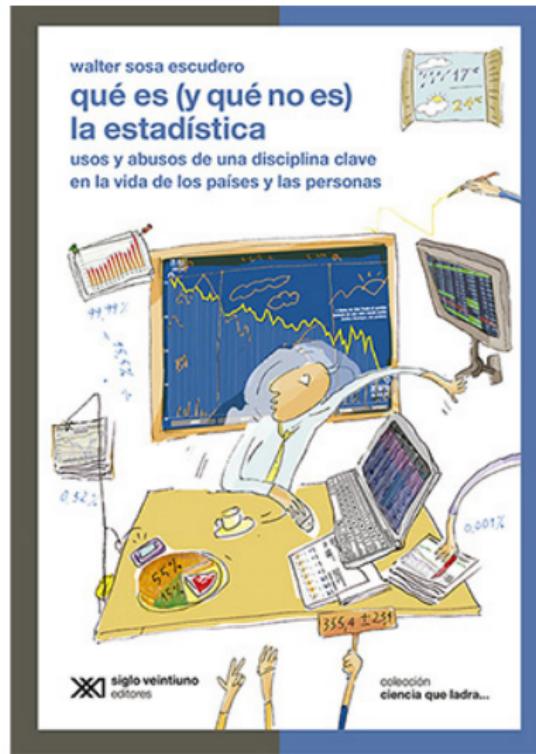
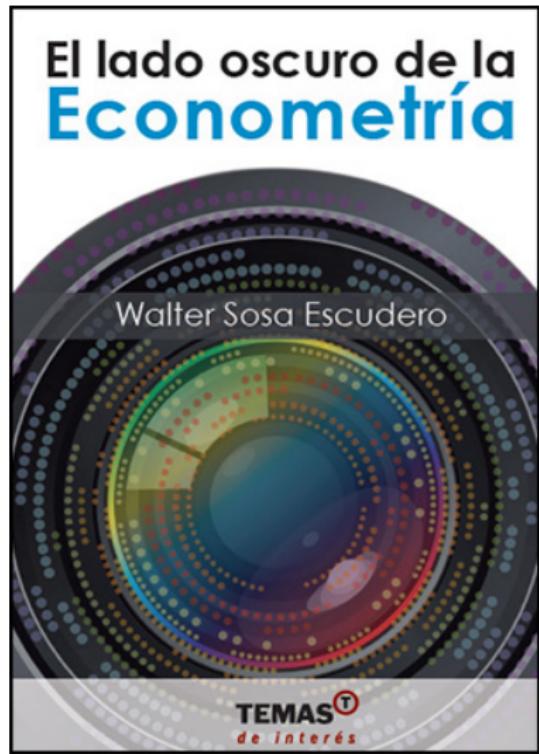
01 NOV 2015



Una gran variedad de cálculos caprichosos son utilizados para forzar resultados de trabajos académicos; la ciencia de hacer que los números digan lo que el autor quiere

Por Walter Sosa Escudero







Home Material Tutoriales Trabajos Prácticos Profesores

Big Data y Minería de Datos en UdeSA

Perspectivas, ideas y herramientas para economistas

Walter Sosa Escudero
Asistente: Noelia Romero

[PROGRAMA](#)

A wide-angle photograph of a sunset or a fire in the sky, with orange and red hues across the horizon and wispy clouds in the foreground.

Referencias

- **Survey I:** Varian, H. R., 2014, Big data: New tricks for econometrics, Journal of Economic Perspectives, 28(2), 3-27.
- **Survey II:** Einav, L., and Levin, J., 2014, Economics in the age of big data. Science, 346(6210).
- **Libro simple:** James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). An introduction to statistical learning. New York: Springer.
- **Libro:** Murphy, K., 2012, Machine Learning: a Probabilistic Perspective, , MIT Press, Cambridge.
- **Predicciones:** Blumenstock, J., Cadamuro, G., and On, R., 2015, Predicting poverty and wealth from mobile phone metadata. Science, 350(6264), 1073-1076.

- **Mediciones:** Cavallo, A. and Rigobon, R. 2016, The Billion Prices Project: Using Online Prices for Measurement and Research, *Journal of Economic Perspectives*, Vol. 30, 2, pp.151-78.
- **Causal:** Athey, S., and G. Imbens (2015), Machine Learning Methods for Estimating Heterogeneous Causal Effects, NBER working paper.
- **Política:** Kleinberg, J., Ludwig, J., Mullainathan, S., and Obermeyer, Z., 2015, Prediction policy problems. *The American Economic Review*, 105(5), 491-495.
- **Econometria:** Belloni, A., V. Chernozhukov, and C. Hansen, 2014, High-Dimensional Methods and Inference on Structural and Treatment Effects, *Journal of Economic Perspectives*, 28(2): 29-50.
- **Proyecto:** Statistical and machine learning for econometricians, Sosa Escudero, W. y Sarmiento, I. (2017).