



UNAM
POSGRADO



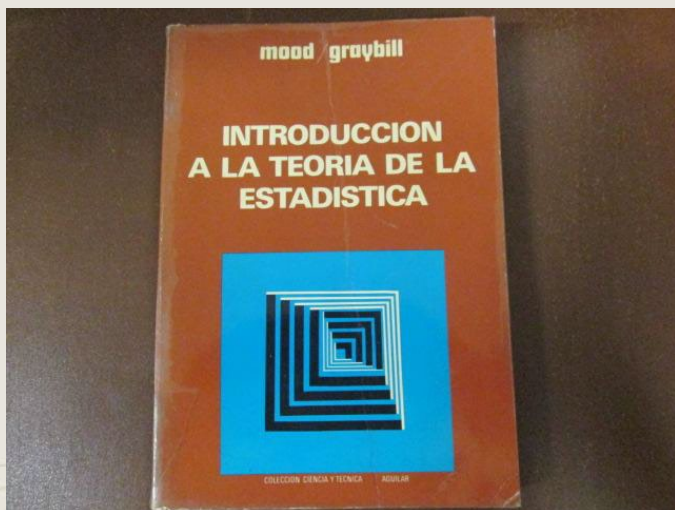
Programa
Universitario
de Estudios
del Desarrollo
UNAM

Presentación del curso

Introducción al análisis de datos bayesiano

Dr. Héctor Nájera
Dr. Curtis Huffman

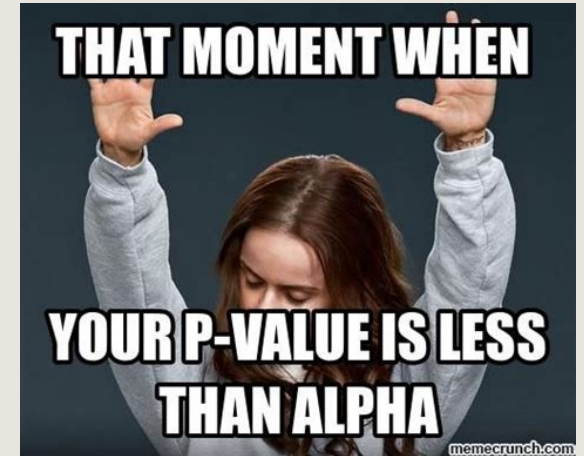
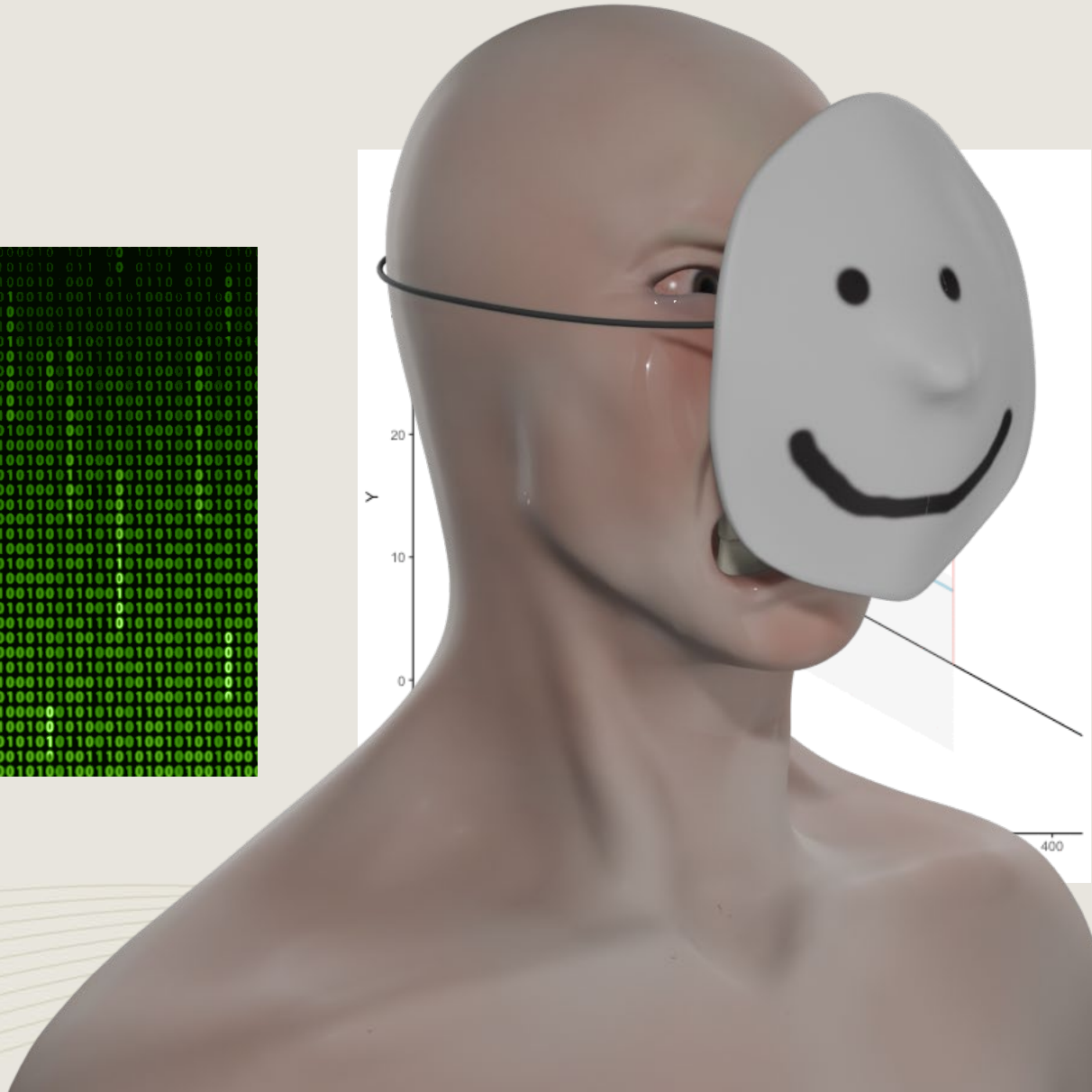
“El nombre completo de la teoría estadística es la teoría de la estadística aplicada”



¿Cómo pasar un proceso de razonamiento que va de unas ideas vagas a conclusiones que permiten tomar cierta decisión?

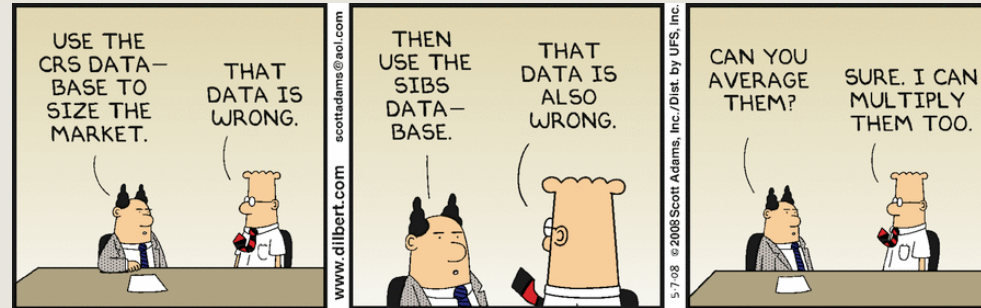


Búsqueda de inferencias válidas



- What's important about the statistical model is not what it does with the data, but what data it uses (Andrew Gelman)

Hacer inferencias válidas es difícil



Datos

Información adicional sobre los datos

Error de medición

Múltiples opciones de modelación

Relaciones causales

Trabajos previos

¿Cómo puedo evitar desordenarme?

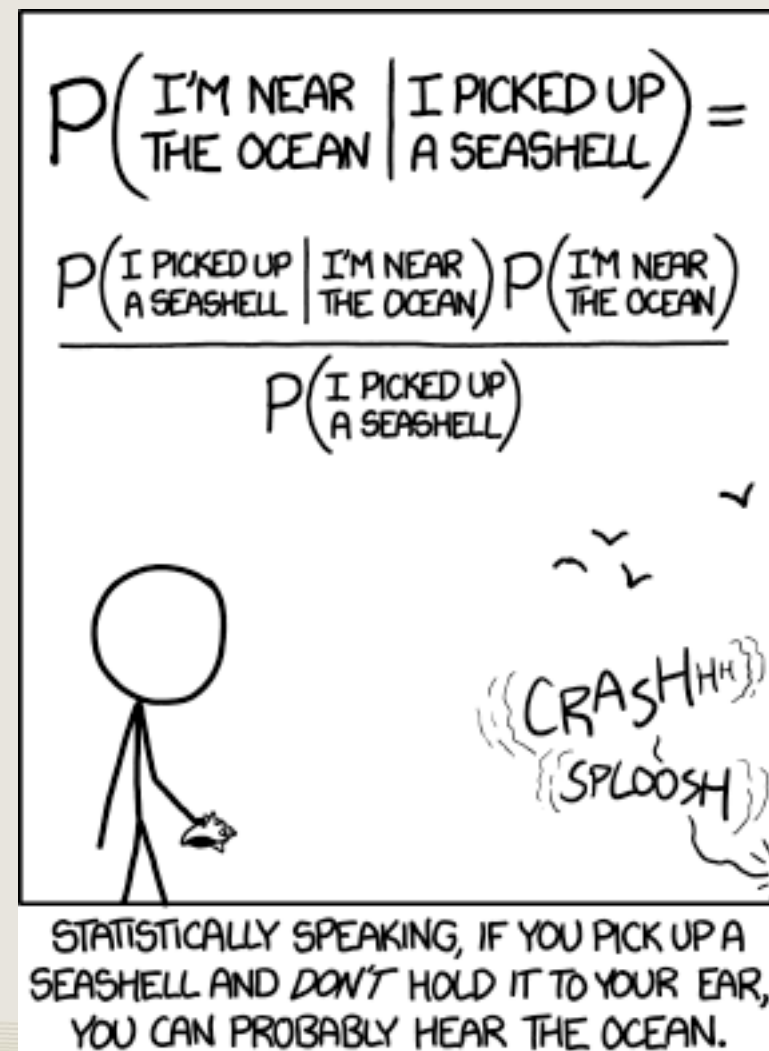
¿Cómo puedo saber que estoy tomando las mejores decisiones?

¿Cómo sé que los resultados son creíbles?

¿Qué tan creíbles (probables) son mis resultados?

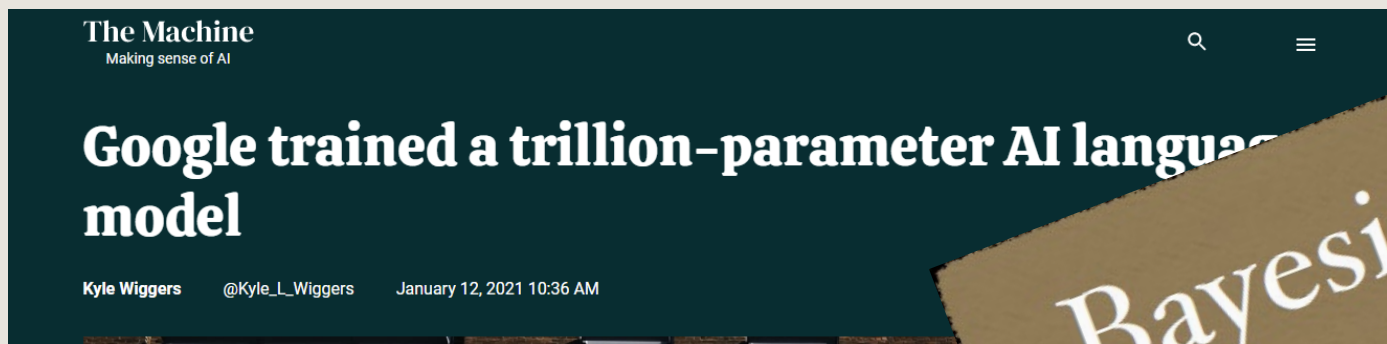
Análisis Bayesiano de datos

O sobre cómo uso las reglas de la probabilidad condicional para ajustar mis conclusiones cuando combino nueva y vieja información



¿Por qué este curso es importante?

- Vivimos en un mundo en el que la investigación empírica en la forma de inferencia estadística está jugando un papel de la mayor importancia, desde el pronóstico de elecciones, el descubrimiento de patrones en los hábitos de consumo (Amazon, Netflix, Facebook), el reconocimiento de voz para el control de cómputo, modelos de lenguaje colosal (LLM), hasta encontrar las bases genéticas de enfermedades como el Covid-19.



Bayesian

Model Name	n_{params}	n_{layers}	d_{model}	n_{heads}	d_{head}
GPT-2 Small	125M	12	768	12	64
GPT-2 Medium	350M	24	1024	16	64
GPT-2 Large	760M	24	1536	16	96
GPT-3 3.5B	1.3B	24	2048	24	128
GPT-3 5.4B	2.7B	32	2560	32	80
GPT-3 175B or "GPT-3"	6.7B	32	4096	32	128
	13.0B	40	5140	40	128
	175.0B	96	12288	96	128



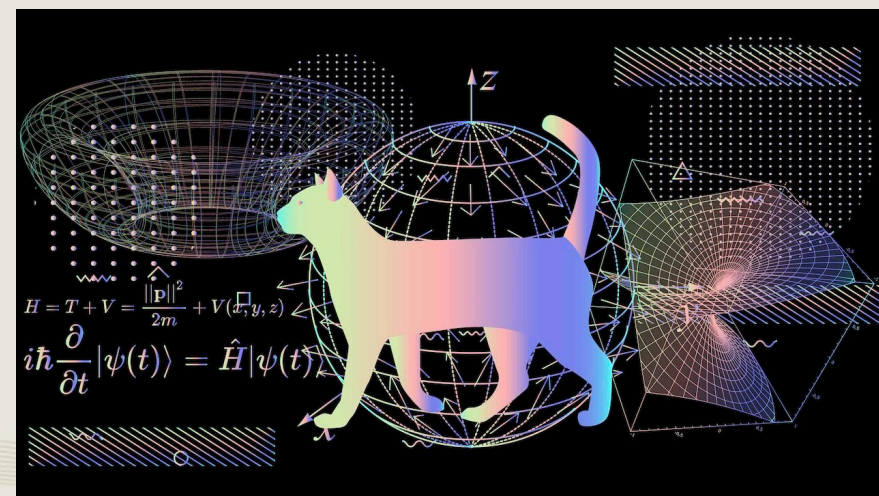
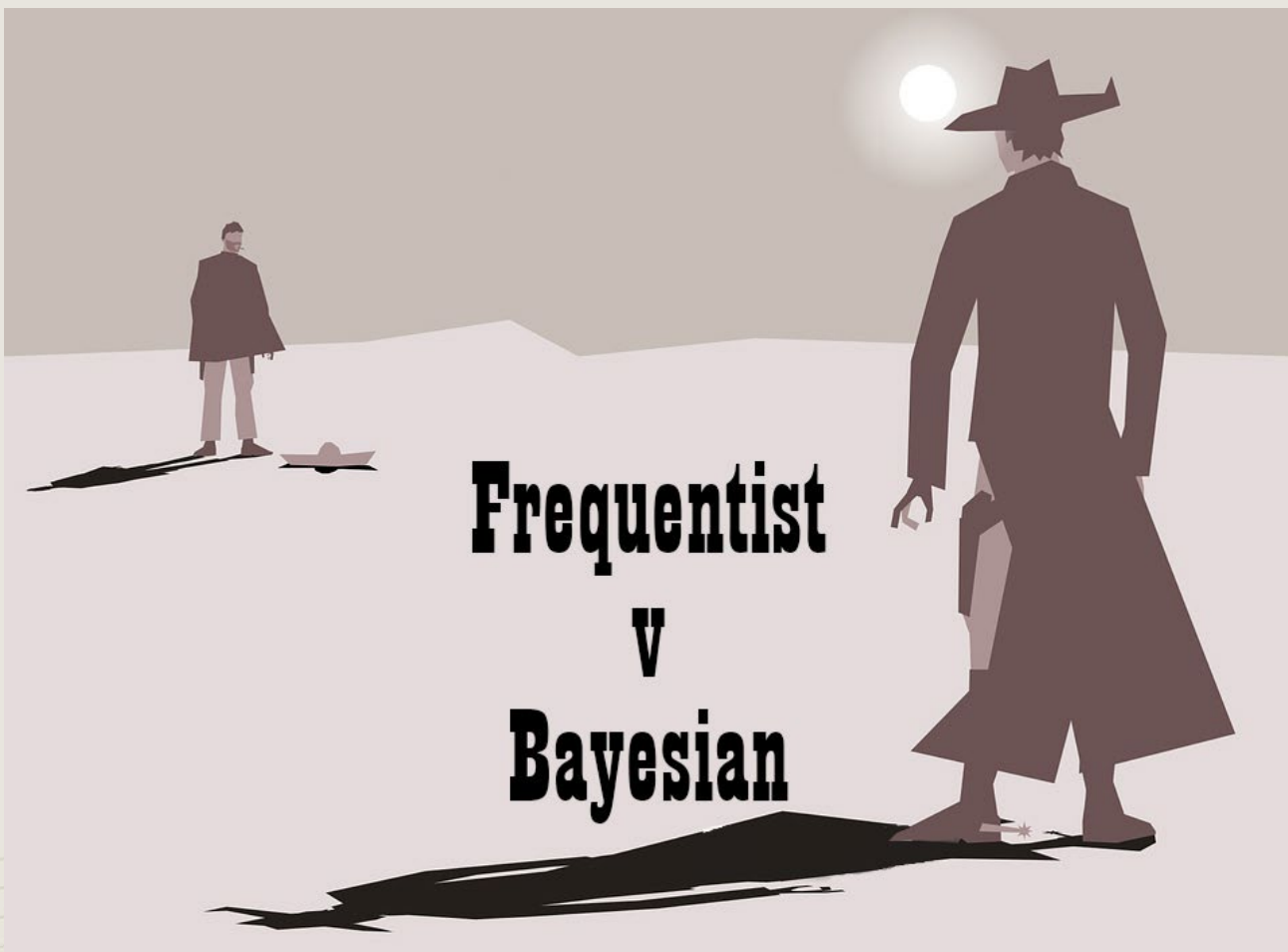


¿Por qué este curso es importante?

- Hoy en día no es posible “andar por la vida” sin “chocar” con la estadística bayesiana
 - se ha convertido en un tema de alfabetismo para la sociedad en su conjunto
 - de cómo interpretar información en la vida cotidiana
 - procesos electorales
 - salud pública
 - y en la vida profesional de la economista
 - cada vez más la profesión requiere un entendimiento cabal de las diferentes maneras en que es posible cuantificar **incertidumbre** usando medidas de probabilidad
 - las fortalezas y debilidades de lo que significa “cuantificar incertidumbre” para diferentes escuelas de estadística (estilos de inferencia)

(Eran) Temas altamente controvertidos

¿Existe la probabilidad/aleatoriedad?





La inferencia bayesiana siempre fue el futuro

- In 2020 most of the meaningful statistical research will be Bayesian (2000) (Donald Rubin)
- A Bayesian statistician is at least as good as a classical one (Gelfan, 2017)
- It is probably better to realize that the probability concept is in a sense *subjective*, that it is always based on uncertain knowledge, and that its quantitative evaluation is **subject to change as we obtain more information**. (Richard Feynman)

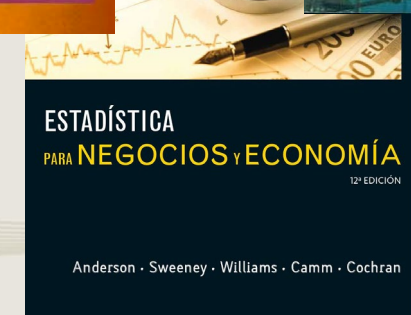
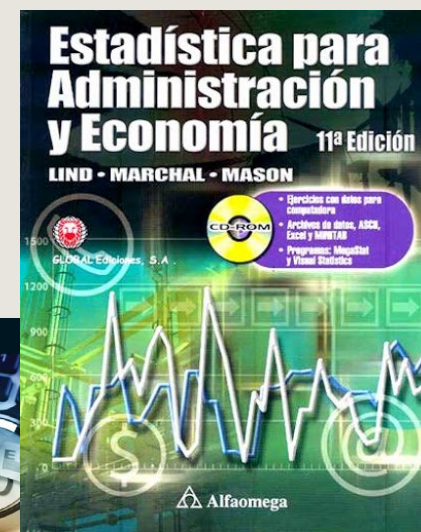
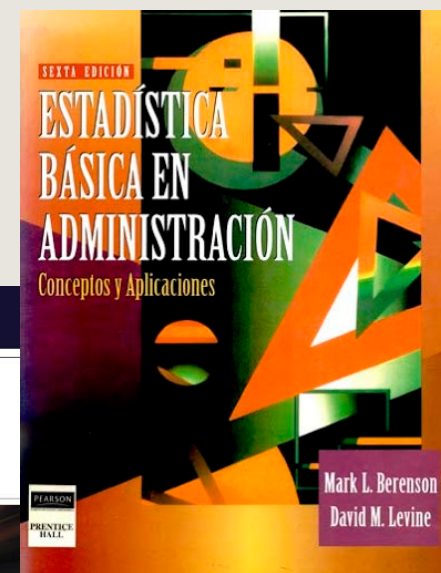
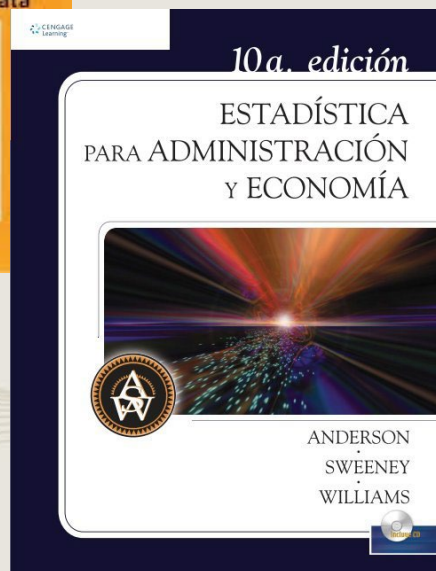
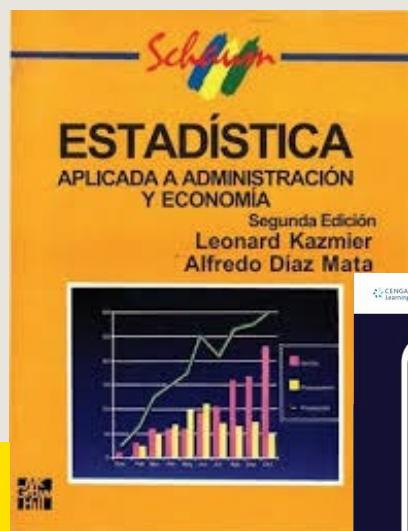
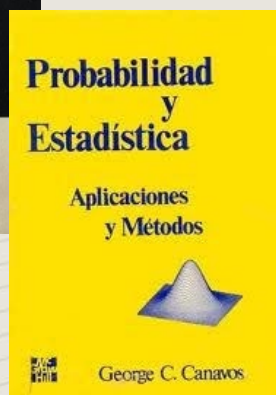


Análisis bayesiano e investigación empírica

- La mayoría de los estadísticos más destacados son Bayesianos
- Las conferencias de punta en estadística son Bayesianas
- Los trabajos de punta (Tesla, Google, universidades top, bolsa de valores, cambio climático, NASA, MLB, NBA, Fórmula 1) se hacen con estadística Bayesiana
- Los programas de estadística en las universidades top en Estados Unidos en su mayoría Bayesianos
- Los modelos más complejos y exitosos que conocemos tienen generalmente un rasgo Bayesiano
- La escritura de nuevo software está pensada para hacer inferencia Bayesiana antes que clásica

Introducción a la estadística

- A pesar de la aplicación exitosa a problemas actuales, existe una importante desconexión entre las mejores prácticas en la investigación empírica y la enseñanza típica a nivel pregrado en México (y algunos posgrados rezagados).



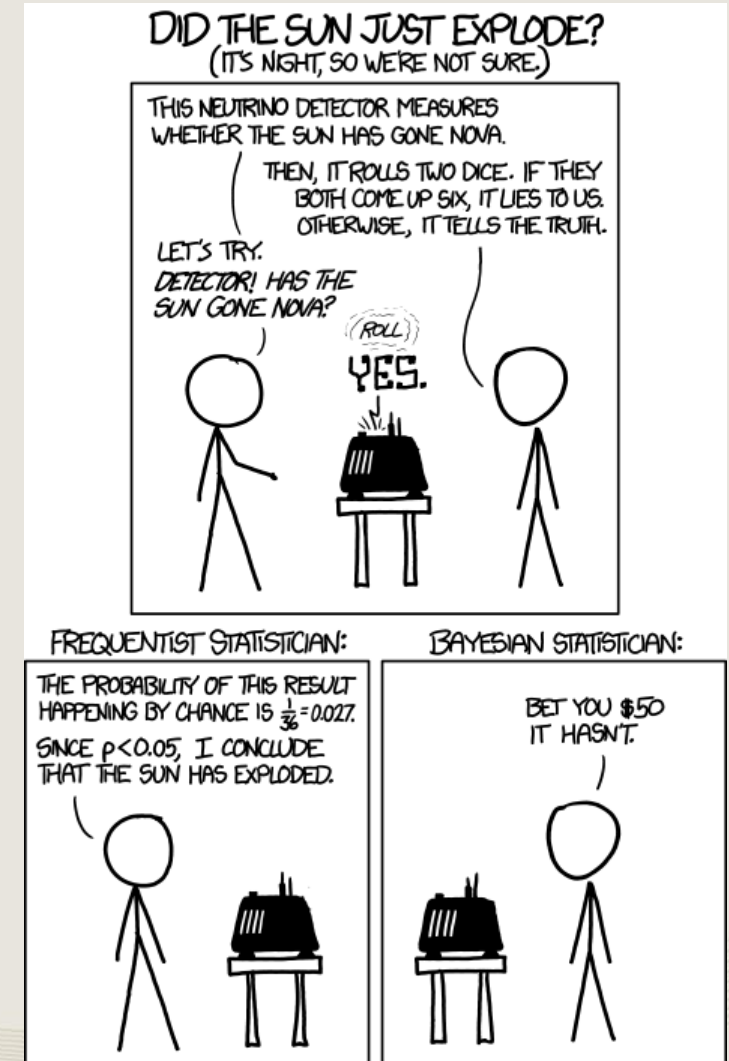


Y es que... las cosas no siempre fueron así

- Aunque muchos de los conceptos clave que usamos hoy en día fueron desarrollados en los siglos XVII y XVIII
 - Esperanza matemática (Huygens 1657)
 - Pruebas de significancia (Arbuthnot 1711)
 - Aproximación a la binomial por la distribución normal (de Moivre 1718)
- Muchos de los primeros métodos estadísticos fueron desarrollados en la última parte del siglo XIX
 - Regresión lineal (Galton 1889)
 - Mínimos cuadrados (Legendre 1805, Gauss 1809)
 - Correlación estadística (Galton 1888, Edgeworth 1893, Yule 1897, Pearson 1896)
 - Medidas de bondad de ajuste (Pearson 1900)
- Pero el campo levantó en serio en la primera mitad del siglo XX
 - Estimación por máxima verosimilitud, ensayos aleatorios (Fisher 1920, 1930)
 - Bases frecuentistas de la prueba de hipótesis (Neyman y Pearson 1933)
 - ANOVA (Yates y Cochran 1938)

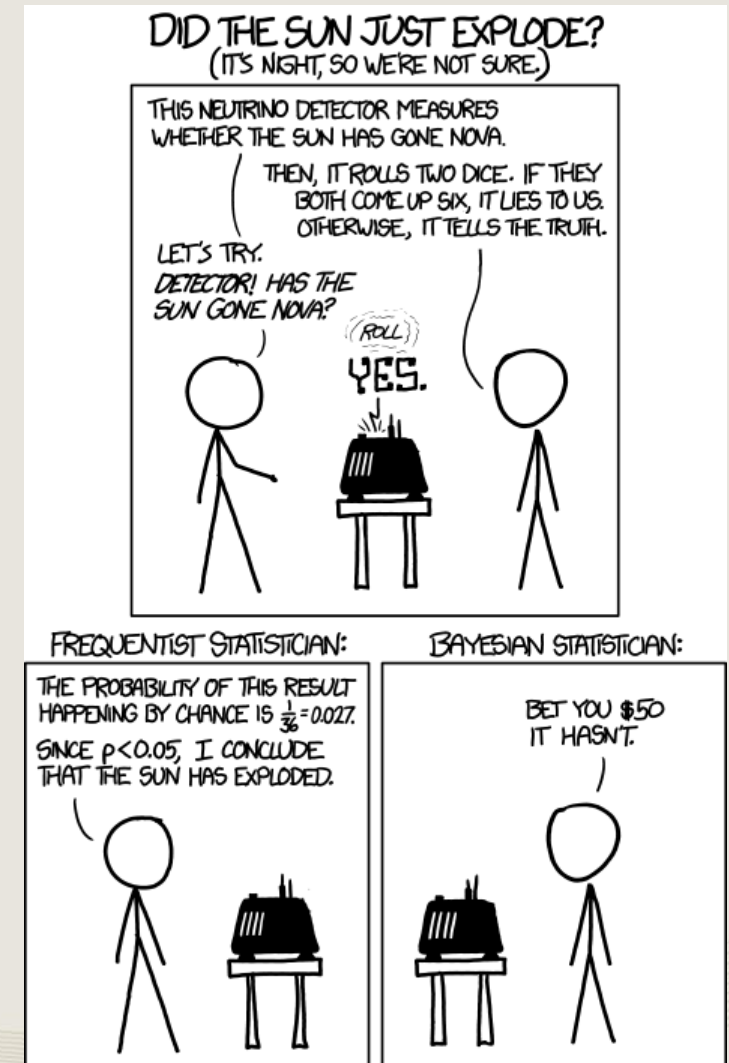
Y es que... las cosas no siempre fueron así

- Aunque los métodos bayesianos se originan en el siglo XVII, en el siglo XX los estadísticos bayesianos eran raros, excéntricos y subocupados
- Dominio de la estadística **clásica o frecuentista** en el siglo XX
 - Los datos dado β (un parámetro genera la realidad)
 - p-values,
 - errores estándar que no indican incertidumbre sino tamaño de muestra
 - MCO y Máxima verosimilitud



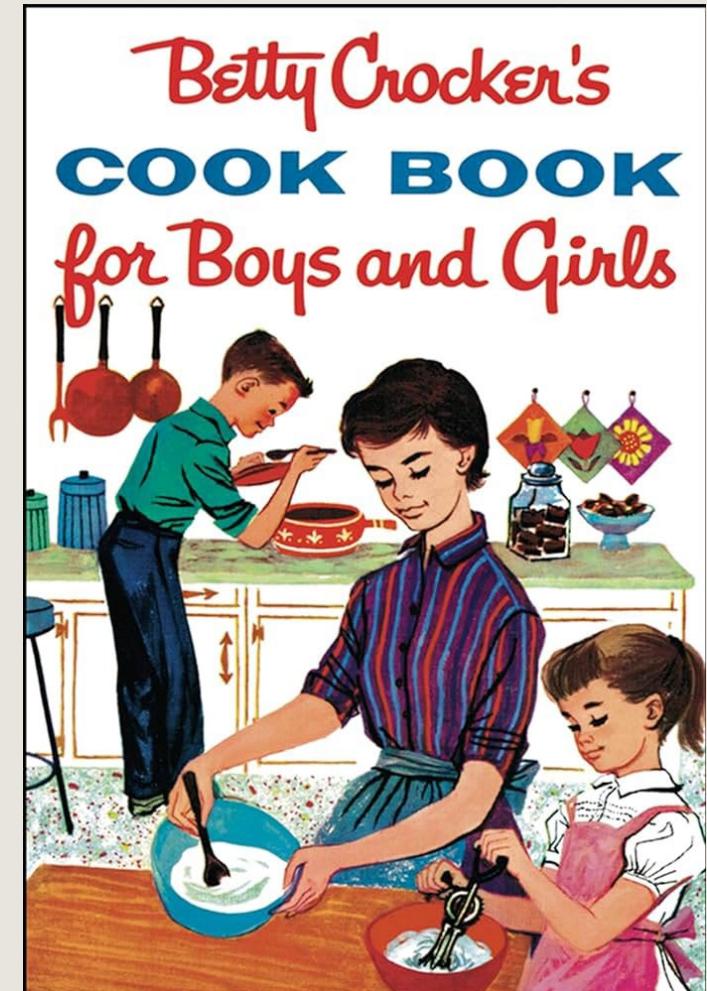
Y es que... las cosas no siempre fueron así

- *Fisherian statistics* (Por Fisher) es “objetiva” e inapelable: descubre la verdad.
- La inferencia basada en p-values e intervalos de confianza es válida y razonable
- No hay razones para desconfiar de la estadística clásica
- La estadística Bayesiana en tanto no es objetiva lleva a soluciones singulares (es necesario pensar en clave de estadística, priors)
- El software estaba pensado a partir de la estadística clásica (reproducible, aunque no sepas que significa)



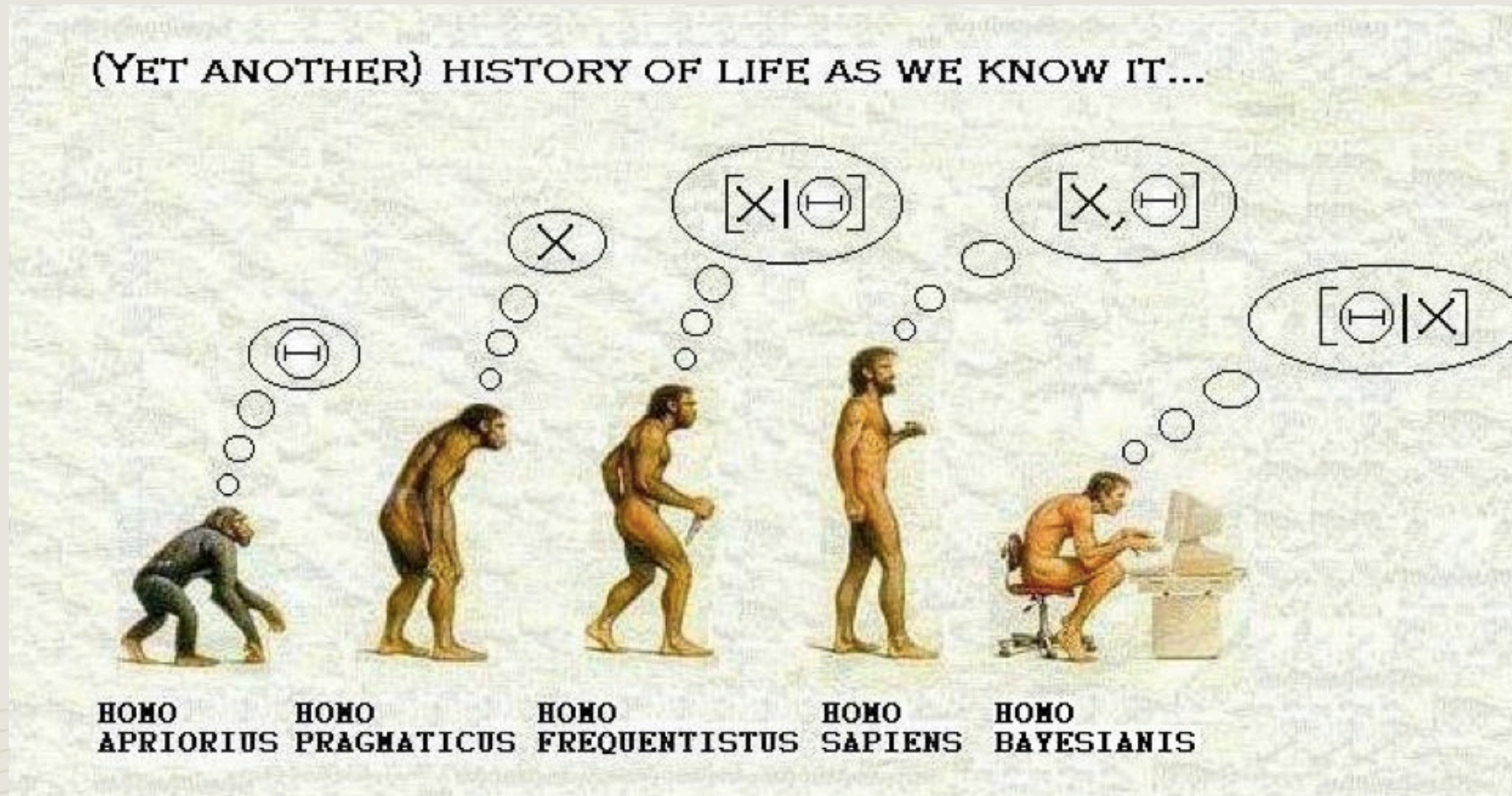
Y es que... las cosas no siempre fueron así

- *Fisherian statistics* (Por Fisher) es “objetiva” e inapelable: descubre la verdad.
- La inferencia basada en p-values e intervalos de confianza es válida y razonable
- No hay razones para desconfiar de la estadística clásica
- La estadística Bayesiana en tanto no es objetiva lleva a soluciones singulares (es necesario pensar en clave de estadística, priors)
- El software estaba pensado a partir de la estadística clásica (reproducibile, aunque no sepas que significa)

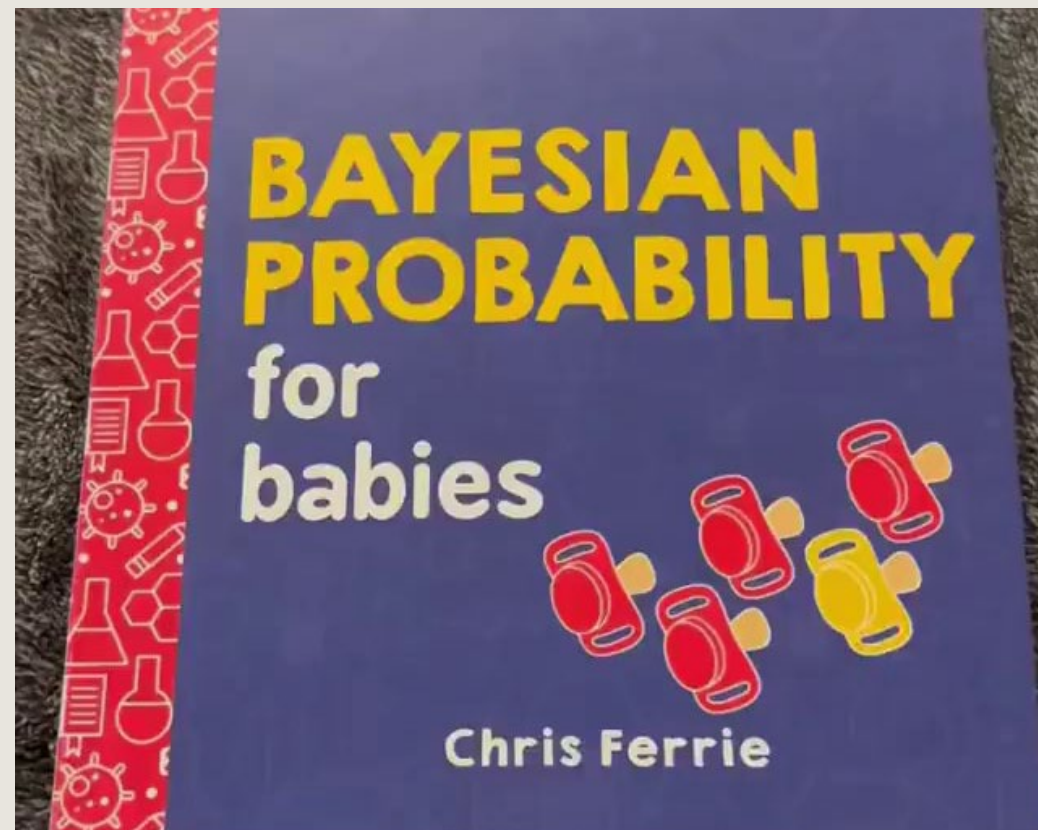


Análisis bayesiano

– ¿Qué originó esta revolución/evolución?



¿Qué es eso del análisis bayesiano y por qué me vengo enterando hasta ahora?



¿Qué es la estadística bayesiana?

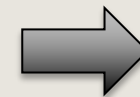
- La estadística bayesiana usa las reglas de la probabilidad y combina toda la información disponible (datos/muestra, conocimiento previo, incertidumbre) para producir inferencias más precisas respecto al caso del uso de información por separado
- En contraste, la estadística clásica evita usar información extra y se concentra en los datos. Puedes incluir en tu modelo un predictor, excluirlo o combinarlo con otros para crear una estimación más estable. Pero esto es lo único que puedes hacer.
- Puedes reproducir los métodos clásicos con inferencia bayesiana, pero con análisis bayesiano puedes hacer mucho más.

¿De dónde viene?

La teoría de la inferencia bayesiana se origina con el fraile Thomas Bayes.

En realidad el primer “verdadero” Bayesiano fue el matemático francés Pierre-Simon Laplace.

Laplace, P. S. (2012). *A philosophical essay on probabilities*. Courier Corporation.

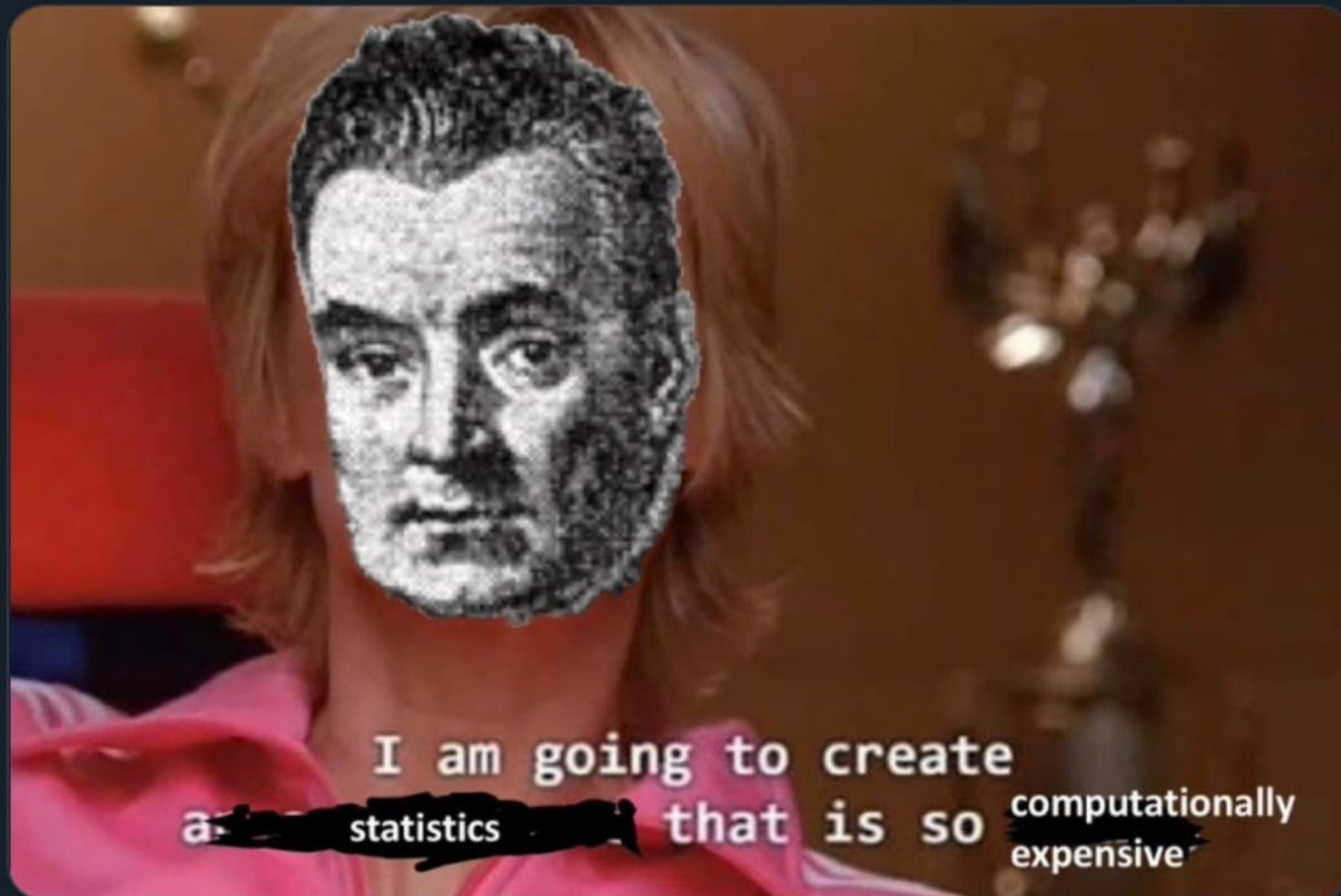


Bayesian in
the flesh!

¿De dónde viene?

- Los métodos Bayesianos fueron usados tiempo después para resolver problemas específicos en ciencias, pero fue hasta la mitad del siglo XX que se propusieron como un marco general estadístico
- Algunas figuras clave fueron John Maynard Keynes y Frank Ramsey, quienes en los 1920s desarrollaron la teoría axiomática de la probabilidad
- Harold Jeffreys and Edwin Jaynes desarrollaron varios métodos para distintos problemas en la física
- Jimmie Savage and Dennis Lindley conectaron su investigación sobre inferencia Bayesiana en los 1950 y 1970s con la estadística clásica
- Alan Turing uso métodos probabilísticos Bayesianos para resolver el código enigma en la segunda guerra mundial

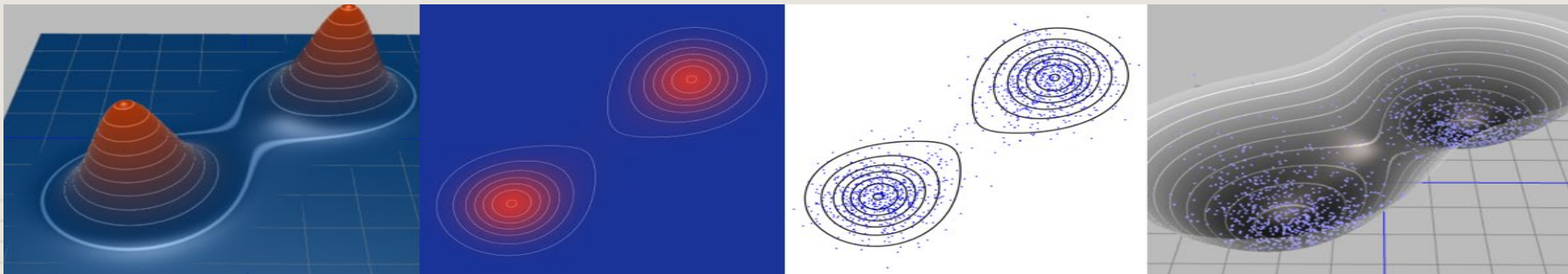
Estadística bayesiana y cómputo



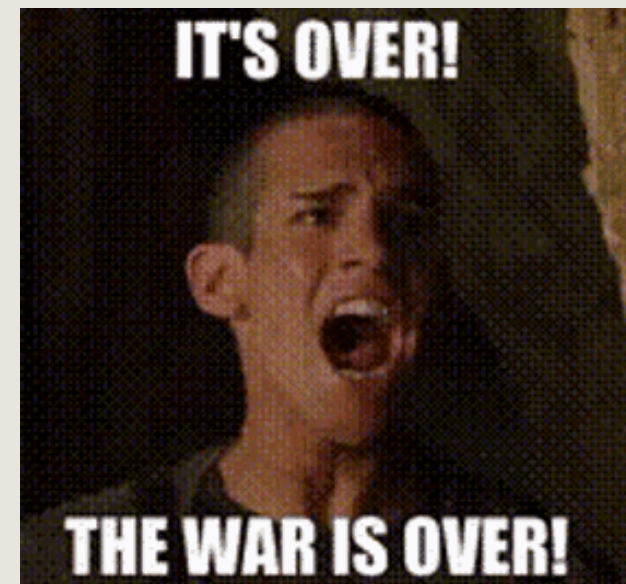
Modelos bayesianos siglo XX



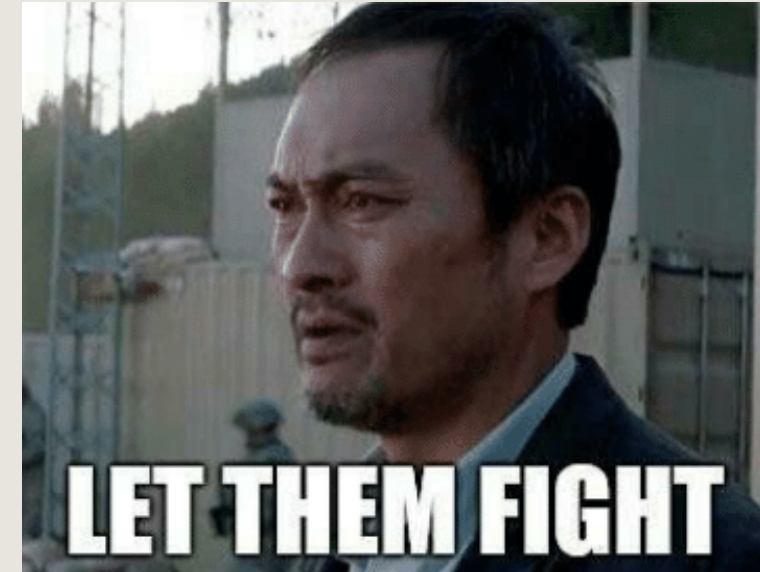
- Un factor clave que permitió la expansión de la inferencia bayesiana a finales del siglo XX fueron las mejoras en la infraestructura para el cómputo y el desarrollo de nuevos algoritmos
- Con excepción de problemas simples, la inferencia Bayesiana requiere cálculos complejos que implican problemas de altas dimensiones e integración numérica.
- La computación bayesiana usa métodos estocásticos de simulación. Esto se conoce como el método de Monte Carlo, que fue desarrollado por Stanislaw Ulam y colegas en los 1940s
- El potencial de estos métodos para resolver problemas irresolubles en estadística fue claro en los 80s. Desde entonces cada década hay saltos mayores en la sofisticación y eficiencia de los algoritmos existentes para resolver preguntas cada vez más complejas



Las guerras de las estadísticas



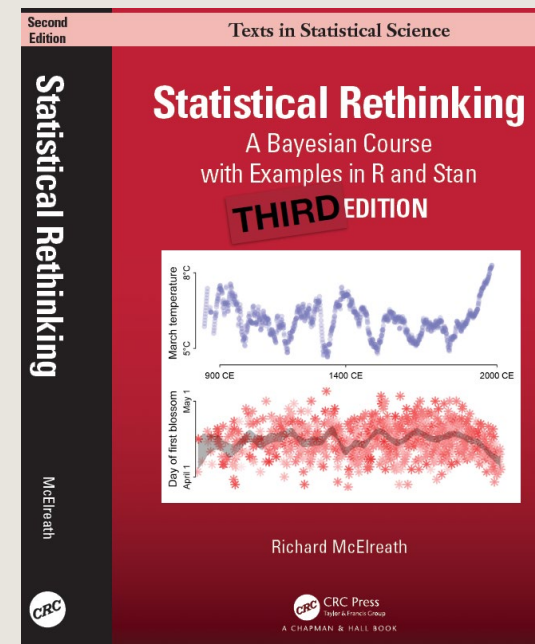
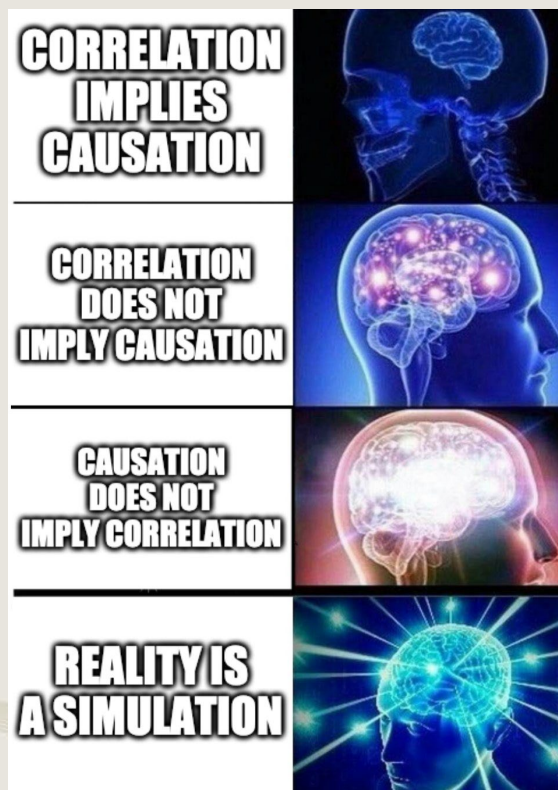
Ciencia antes que estadística



Estadística aplicada a la investigación empírica: conectar datos de nuestro objeto de estudio con los modelos científicos que explican las regularidades de su evolución

Ciencia antes que estadística

Las razones detrás del análisis estadístico no se encuentran en los datos mismos, sino en las causas de los datos





Somos más bayesianos que clásicos

- Considere $\beta = 4$ IC 95 [2-6]. Entonces $p\text{-value} < .05$
- La esencia de la estadística Bayesiana es la combinación de información de distintas fuentes. Esto se conoce como “prior information” o “información a priori”, o modelación jerárquica, actualización dinámica, o agrupación parcial.
- La estadística bayesiana junta todos los datos (información) para entender una estructura superior
- Los datos son muchos pero vienen de fuentes con error (muestreo y medición) y se necesita un marco que ponga a la inferencia en un mismo sitio.

Algunos ejemplos

- Elecciones en USA (2020)
- Algoritmos de reconocimiento de fotografías
- Estimación de áreas pequeñas
- Las parejas *guapas* tienen más hijas...
 - ¡No! De hecho Bayes evita pifias de los métodos clásicos basados en p-values.

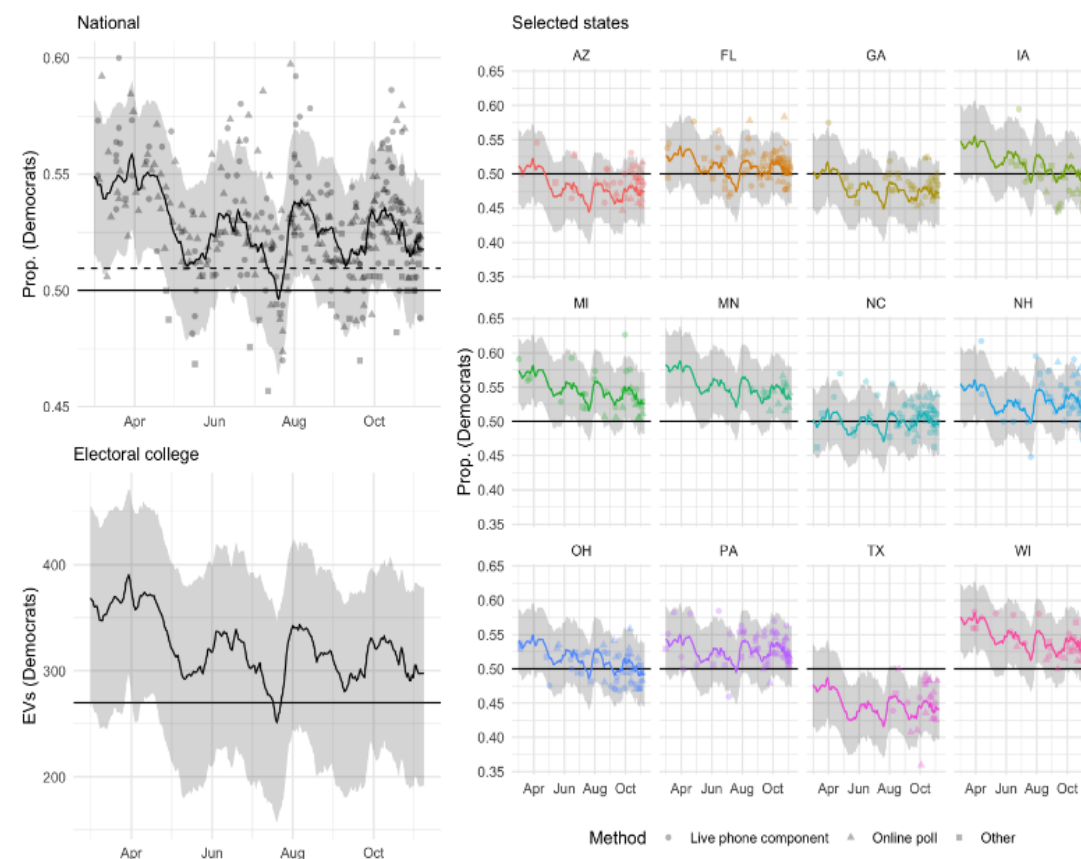


Figure 1: *Some summaries of the model, as fit retrospectively to using state and national polls from 2016. These graphs illustrate that our data and model are fitting national as well as separate state trends.*

Estimación de áreas pequeñas

Error:

Bayes vs
alternativas

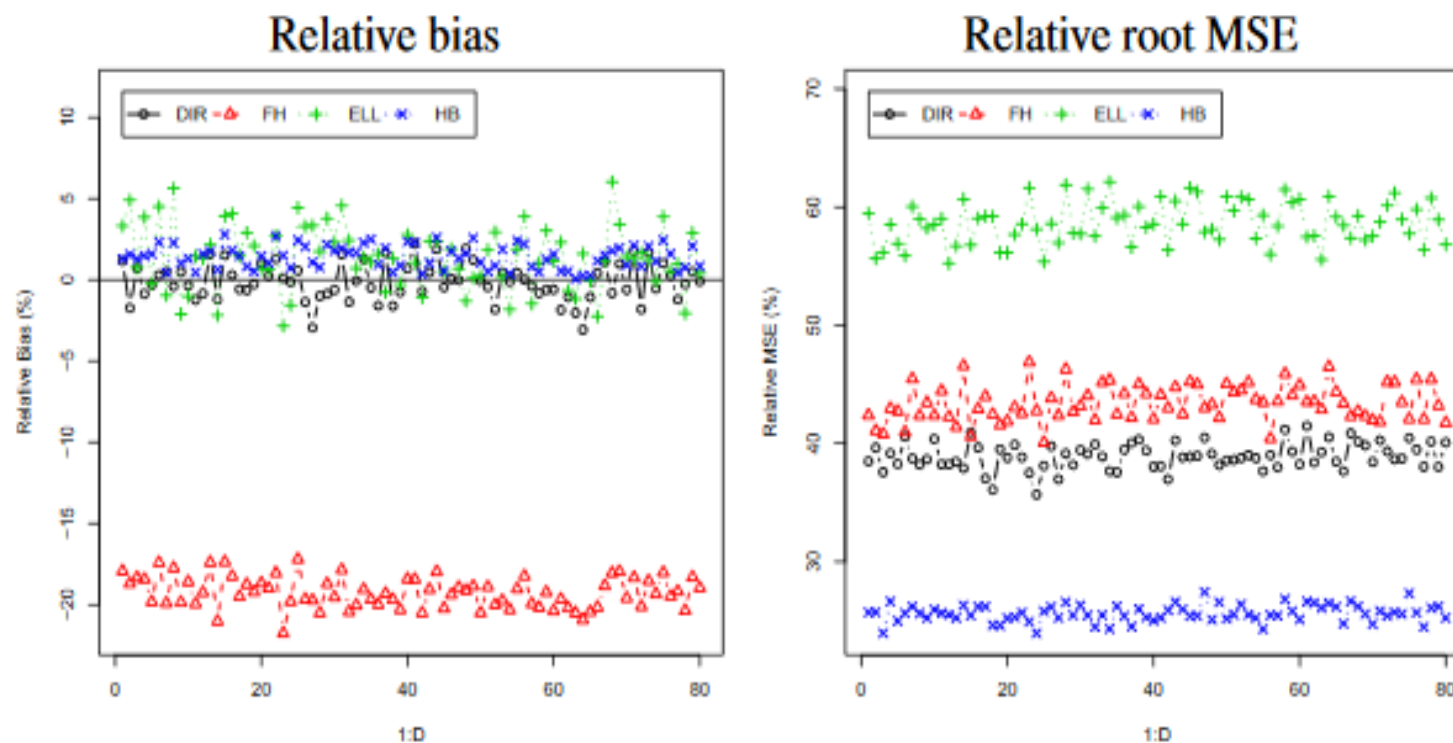


Figure 4. Percent relative bias (left) and relative root MSE (right) of direct, FH, HB and ELL estimators of poverty gap F_{1d} for each area d under low informativeness.

Bayes desbanca en muchos casos

Social Science Research 104 (2022) 102689



Contents lists available at [ScienceDirect](#)

Social Science Research

journal homepage: www.elsevier.com/locate/ssresearch



Bridging the gap between multilevel modeling and economic methods

Aleksey Oshchepkov^{a,*}, Anna Shirokanova^b

^a Centre for Labour Market Studies and Department of Applied Economics, HSE University, Moscow, Russian Federation

^b Ronald F. Inglehart Laboratory for Comparative Social Research and Department of Sociology, HSE University, Saint-Petersburg, Russian Federation

“In sum, MLM generally can do all the things that `economic methods’ should do in a nested data setting and do the even better”



Objetivo

El objetivo del curso es introducir conceptos básicos, métodos y paquetería estadística de punta para llevar a cabo inferencia bayesiana, así como preparar al estudiante con experiencia práctica y nociones metodológicas para aplicar los métodos bayesianos a problemas reales.

Este es un curso de estadística aplicada, no de cálculo diferencial e integral, ni teoría de la medida



Objetivo

El objetivo del curso NO es convertirlos al bayesianismo, sino ayudarles a adquirir la claridad necesaria para distinguir las mejores herramientas para cuantificar incertidumbre cuando así lo requieran en su vida profesional.

- Dado un problema a resolver, si les parece que tiene sentido un abordaje particular, y expresan con claridad sus decisiones de modelaje teórico y estadístico, el curso habrá cumplido su objetivo.

Expectativa

Al final del curso se espera que los alumnos sean capaces de:

- Identificar los usos apropiados e inapropiados de métodos estadísticos
- Aquilatar las prácticas actuales de inferencia estadística y sus resultados
- Entender las ventajas de hacer investigación bajo el paradigma bayesiano
- Interpretar los resultados del análisis bayesiano de una forma crítica
- Reportar apropiadamente los resultados del análisis bayesiano
- Plantear problemas de toma de decisiones bajo incertidumbre en la tradición bayesiana
- Construir sus propios modelos para hacer inferencia probabilística en la tradición bayesiana

Introducción al análisis bayesiano

- Es un curso introductorio sobre las formas y el análisis de datos bajo el esquema de la estadística bayesiana
- Es mucho más aplicado que teórico
- Nos basaremos en el libro de McElreath “Statistical Rethinking”. Second Edition
- Introduciremos la estimación de modelos bayesianos con **R-Software**

MODIFIED BAYES' THEOREM:

$$P(H|X) = P(H) \times \left(1 + P(C) \times \left(\frac{P(X|H)}{P(X)} - 1 \right) \right)$$

H: HYPOTHESIS

X: OBSERVATION

P(H): PRIOR PROBABILITY THAT H IS TRUE

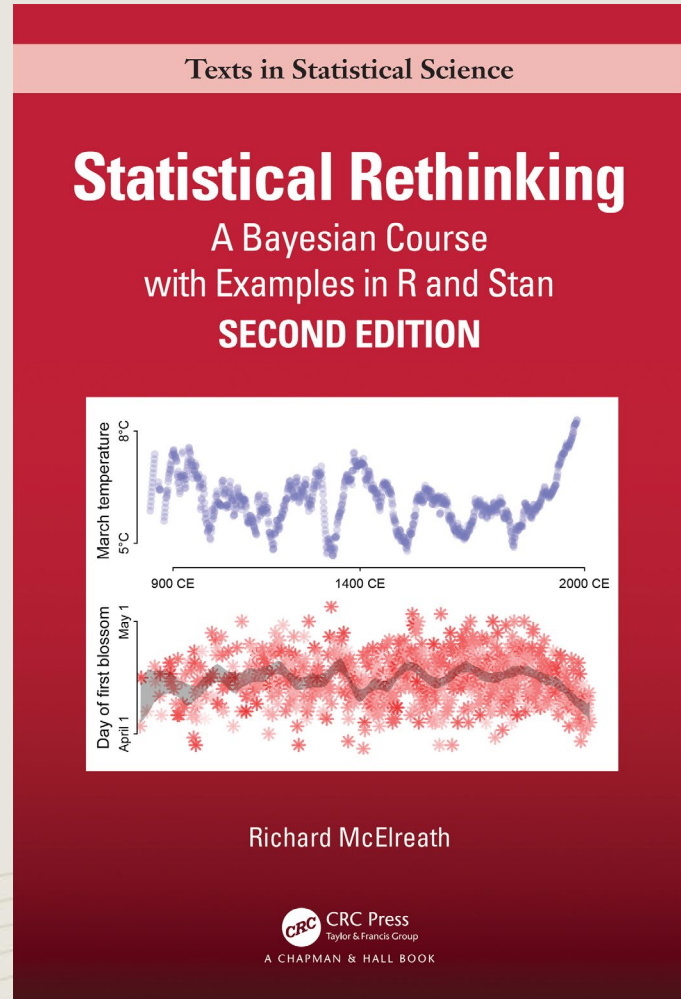
P(X): PRIOR PROBABILITY OF OBSERVING X

P(C): PROBABILITY THAT YOU'RE USING
BAYESIAN STATISTICS CORRECTLY

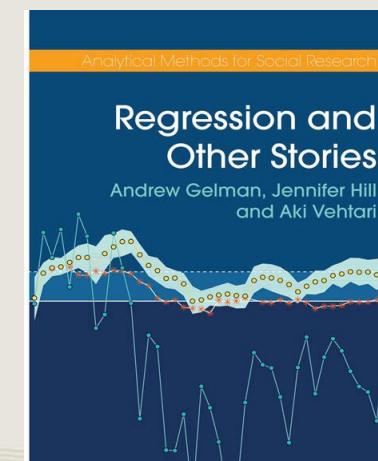
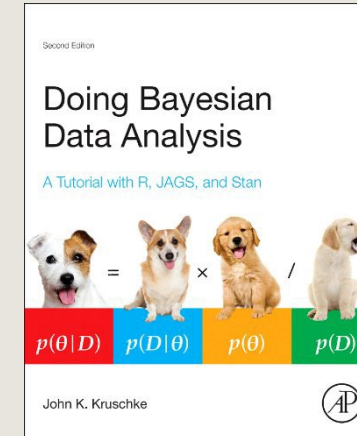
Características de las sesiones

- Las sesiones combinan discusión, teoría y aplicación con el programa R.
- Antes de cada clase, los alumnos deberán leer una selección de capítulos/artículos para su discusión en grupo.
- Los docentes impartirán cada sesión (prepararán un archivo.ppt que subirán a Github después de cada clase) y se dedicará siempre un espacio para discusión, ejercicios en grupo y/o implementación de análisis usando el programa R.

Básica



Secundaria





Evaluación

Se utilizará un ejercicios teórico-prácticos para valorar los contenidos que los alumnos manejan con confianza y aquellos que necesitan reforzarse.



Temario (General)

1. Conceptos básicos: modelos, inferencia, probabilidad, regla de Bayes, y R

Inferencia, Credibilidad, modelos y parámetros; ¿qué es esa cosa llamada probabilidad?; regla de Bayes; el ambiente *R* (*r-project*)

2. Fundamentos aplicados a la inferencia de probabilidades binomiales

Inferir probabilidades binomiales; técnicas de Monte Carlo vía cadenas de Markov; modelos jerárquicos; abordaje bayesiano a las pruebas de significancia de hipótesis nulas; abordajes bayesianos a la prueba de hipótesis puntuales

3. Fundamentos de modelación bayesiana

Combinaciones lineales y predictores; regresión lineal simple, múltiple y logística; calibración bayesiana en modelos de regresión; evaluación bayesiana del modelo de regresión; expansión bayesiana del modelo de regresión; flujo de trabajo bayesiano



Materiales: Github y drive

Los docentes utilizarán esta plataforma para compartir los materiales del curso (bibliografía, presentaciones, ejercicios). La dirección relevante es:

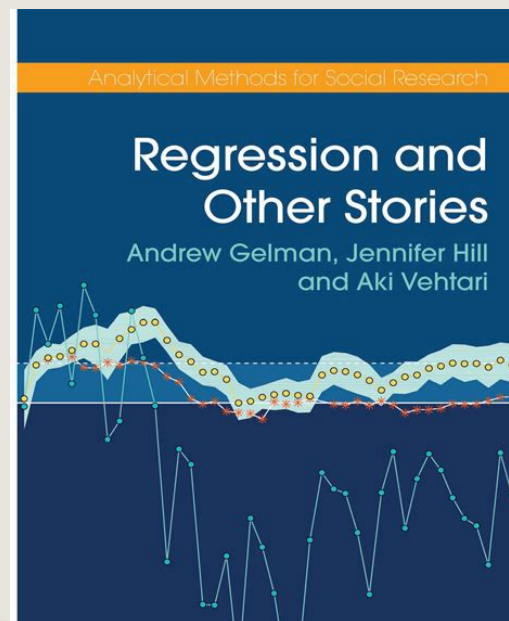
<https://github.com/hectornajera83/ClaseBayes2025/>

Lugar y hora

16 sesiones los jueves de 16:00 a 19:00

Esencial:

Capítulo I. Regression and other stories. Gelman, Hill y Vehtari.



— Recomendada

10.1098/rsta.2003.1263



Bayesian computation: a statistical revolution

BY STEPHEN P. BROOKS

*Statistical Laboratory, University of Cambridge, Centre for Mathematical Sciences,
Wilberforce Road, Cambridge CB3 0WB, UK (steve@statslab.cam.ac.uk)*

Published online 3 November 2003

The 1990s saw a statistical revolution sparked predominantly by the phenomenal advances in computing technology from the early 1980s onwards. These advances enabled the development of powerful new computational tools which reunited



Próxima clase

— Recomendadas

- Lennox, K. [Lawrence Livermore National Laboratory]. (2016, September 27). *All About that Bayes: Probability, Statistics, and the Quest to Quantify Uncertainty* [Video]. YouTube.
<https://youtu.be/eDMGDhyDxuY>
- McElreath, R. [Richard McElreath]. (2023, January 2). Statistical Rethinking 2023 - Lecture 01 *The Golem of Prague* [Video]. YouTube.
<https://youtu.be/FdnMWdICdRs>



Referencias

- Andrew Gelman (2011), “Induction and Deduction in Bayesian Data Analysis“, Special Topic: Statistical Science and Philosophy of Science RMM Vol. 2, 2011, 67–78
- Brooks, S. P. (2003). Bayesian computation: a statistical revolution. Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences, 361(1813), 2681-2697.
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis*. CRC press.
- Hernández, D.R. 2007. *Introducción al análisis bayesiano*. Mar del Plata : Instituto Nacional de Investigación y Desarrollo Pesquero INIDEP. 45 p.
- Kruschke, J. (2014). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan*. Academic Press.
- Laplace, P. S. (2012). *A philosophical essay on probabilities*. Courier Corporation.
- McElreath, R. (2020). *Statistical rethinking: A Bayesian course with examples in R and Stan*. CRC press.



CONTACTO

Dres. Héctor Nájera y Curtis Huffman
Investigadores (SNI-II)

Programa Universitario de Estudios del Desarrollo (PUED)
Universidad Nacional Autónoma de México (UNAM)

Antigua Unidad de Posgrado (costado sur de la Torre II de Humanidades), planta baja.
Campus Central, Ciudad Universitaria, Ciudad de México, México.

Tel. (+52) 55 5623 0222, Ext. 82613 y 82616

Tel. (+52) 55 5622 0889

Email: hecatalan@hotmail.com
chuffman@unam.mx

*¡Bienvenidos
estudiantes!*

