



UNAM
POSGRADO



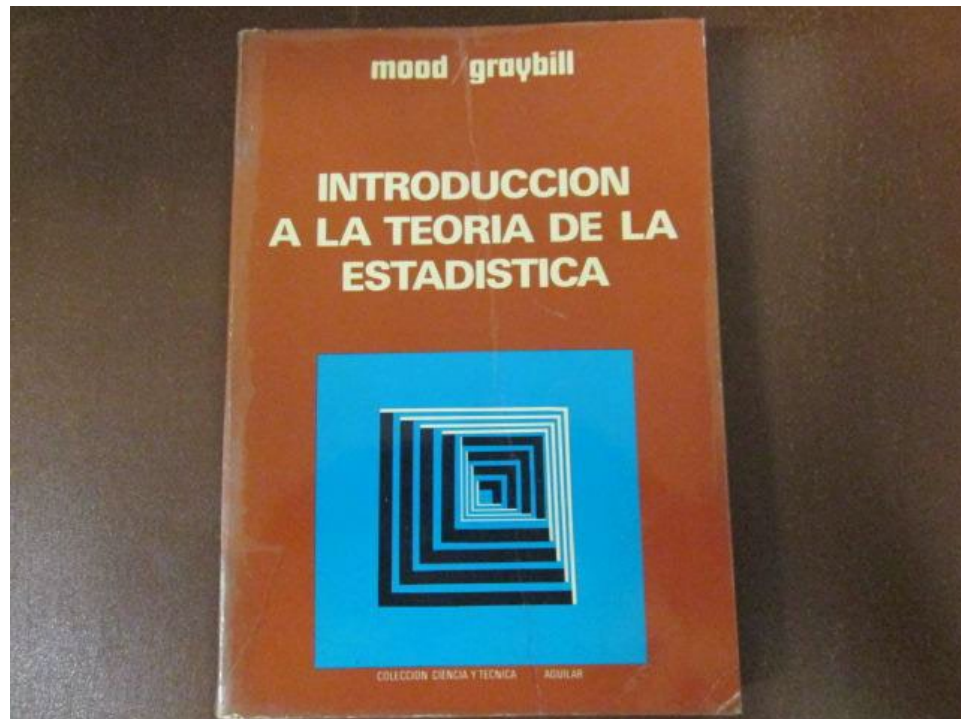
Programa
Universitario
de Estudios
del Desarrollo
UNAM

Introducción al análisis bayesiano de datos

Dr. Curtis Huffman

Dr. Héctor Nájera

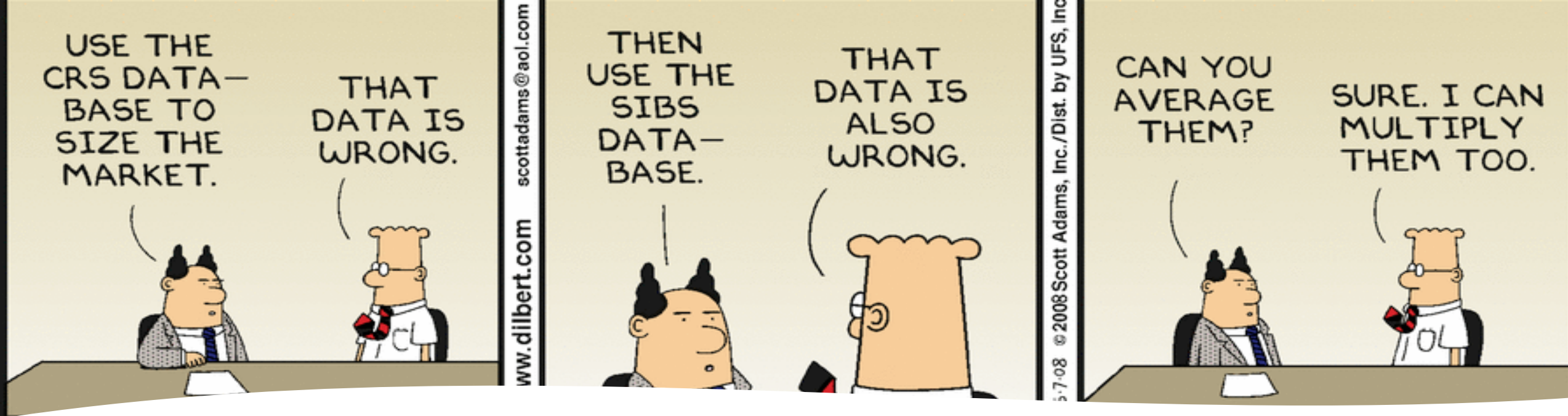
La estadística como disciplina en la economía



*“El nombre completo de la teoría estadística es la teoría de la estadística **aplicada**”*
(Gelman, 2003)

¿Por qué empleamos a la estadística?

- Generalizar:
 - De una muestra a una población
 - De un conjunto de observables a un constructo de interés
 - Del grupo de tratamiento al de control



Ciencia antes que estadística

- La generalización (inferencia) demanda modelos teóricos sobre la población, aquello que queremos medir, y las relaciones causa-efecto de interés.
- Necesitamos un proceso de razonamiento y estimación que conecte satisfactoriamente nuestros intereses de investigación con conclusiones generalizables con *cierto grado de precisión*

Generalizar con precisión es difícil

Modelos teóricos pobremente especificados

Datos no creados para
investigar lo que me interesa

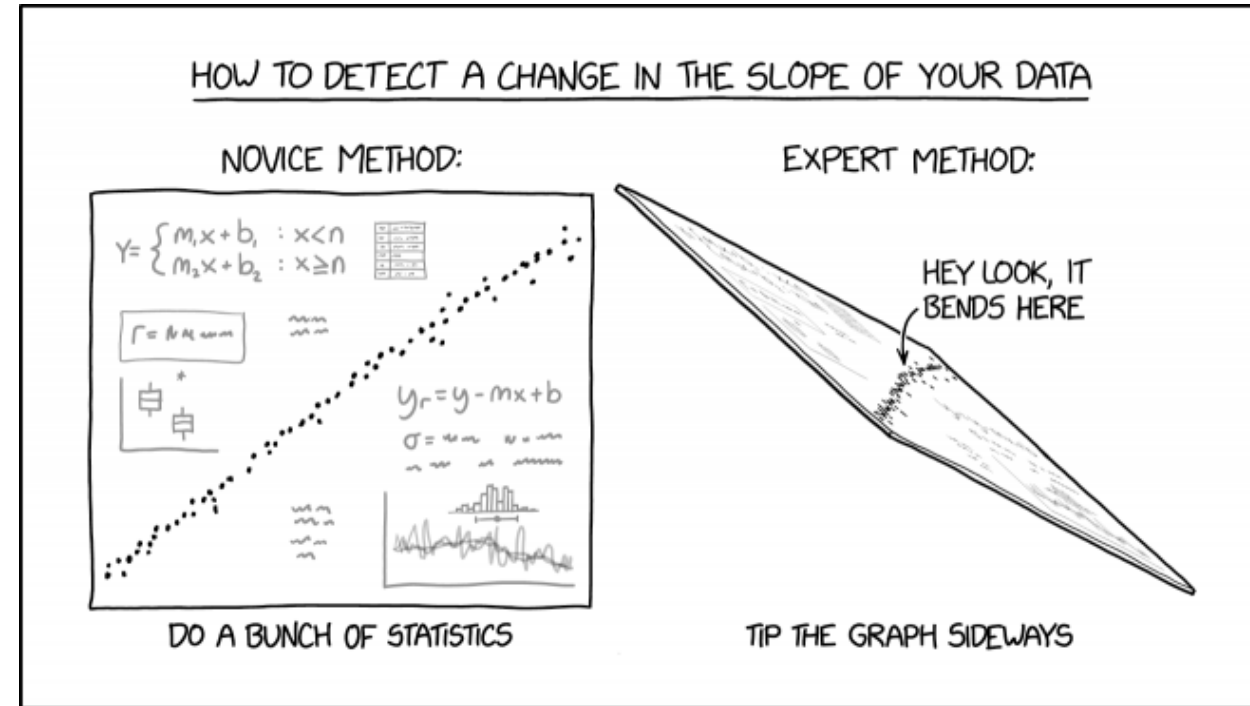
Múltiples opciones de modelación

Información adicional sobre los datos

Error de medición

Relaciones causales

Trabajos previos



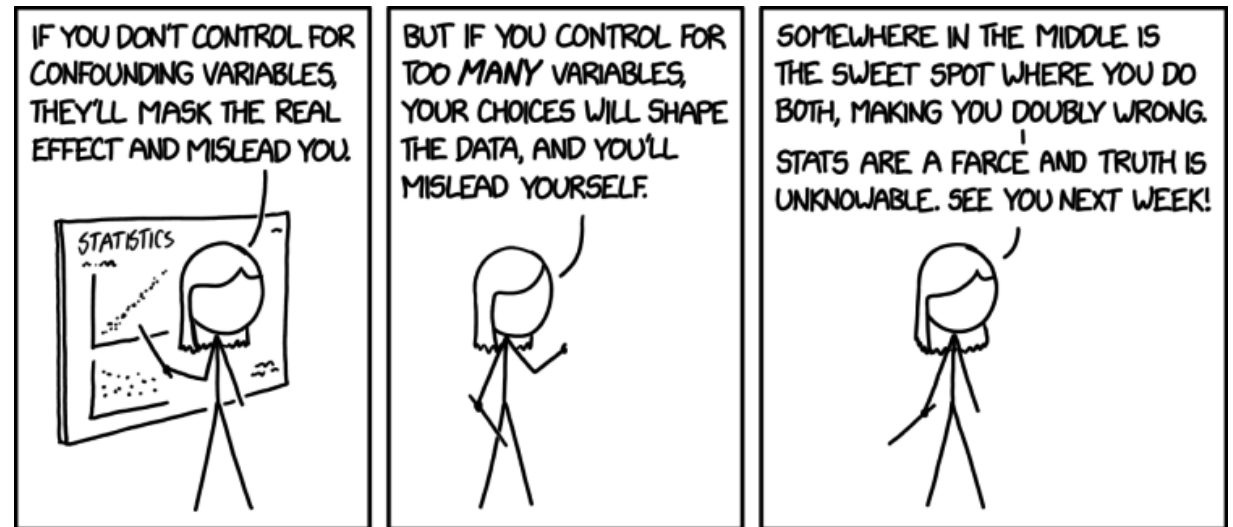
Generalizar con precisión es difícil

¿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?



Repuesta típica de la enseñanza a nivel pregrado en México (y algunos posgrados rezagados).

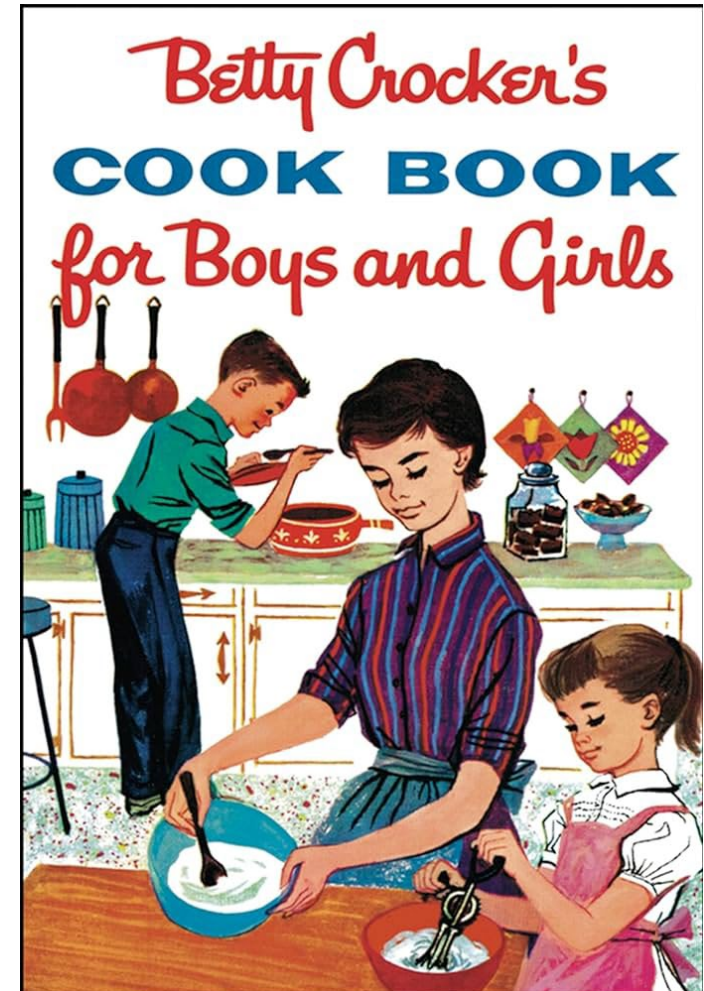


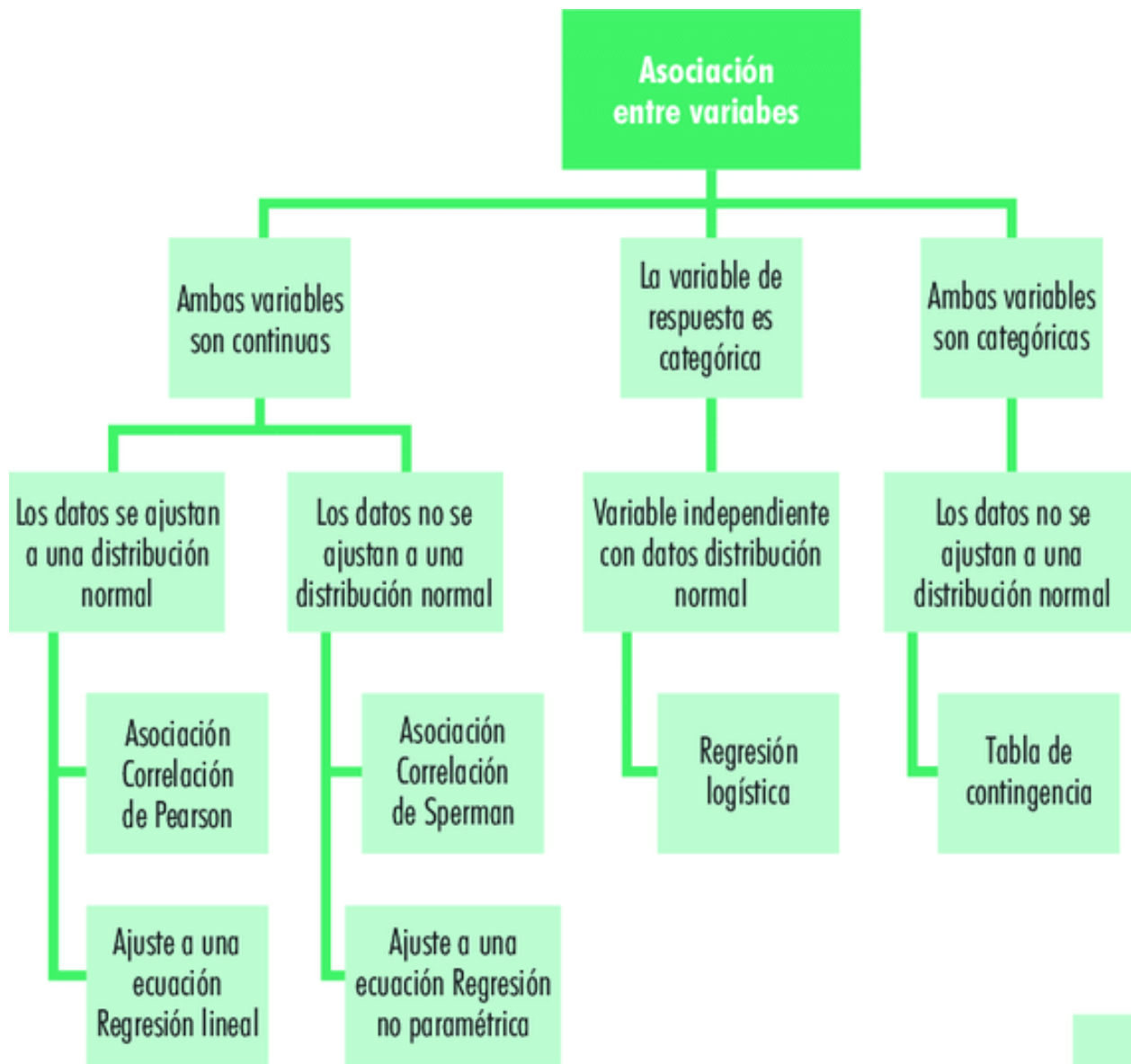
Estadística clásica... simplemente se podía

- 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

- *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
- El software estaba pensado a partir de la estadística clásica (reproducible, aunque no sepas que significa)

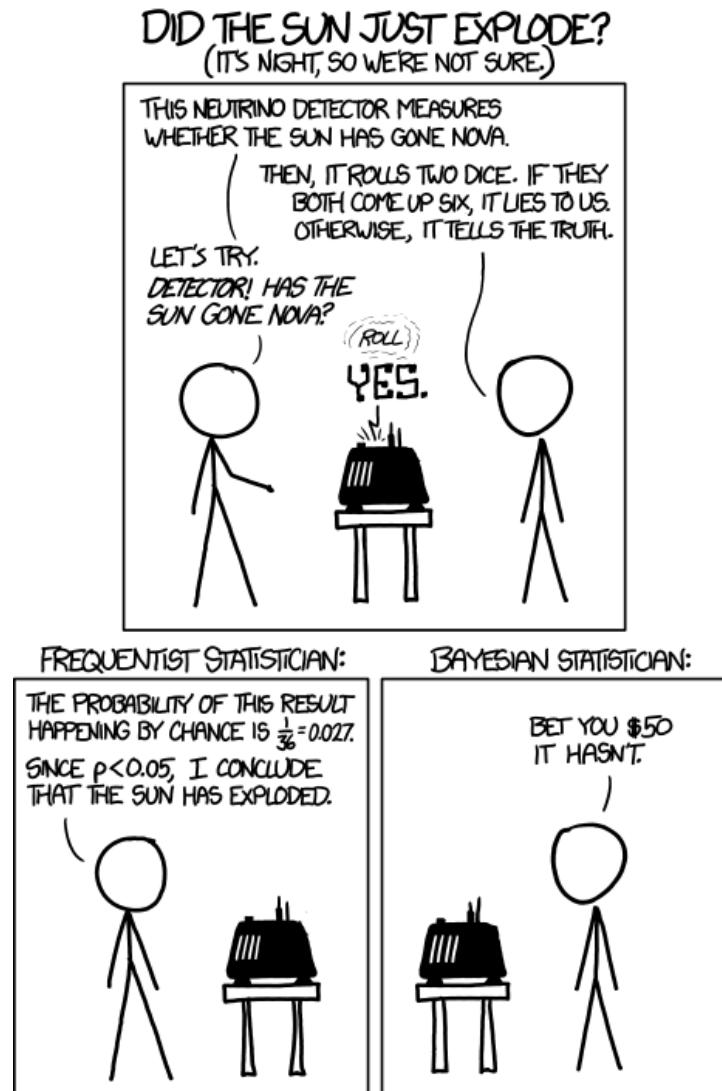




¿Son razonables estos “pasos” para conectar mis intereses de investigación y los problemas típicos del análisis de datos con los modelos estadísticos?



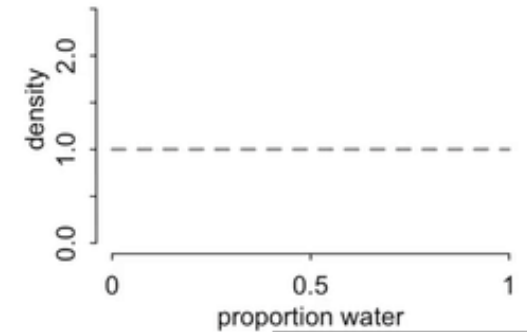
¿Qué tipo de inferencia nos importa?



Siete aspectos comunes del proceso de generalización

(y por qué la inferencia bayesiana es el único camino)

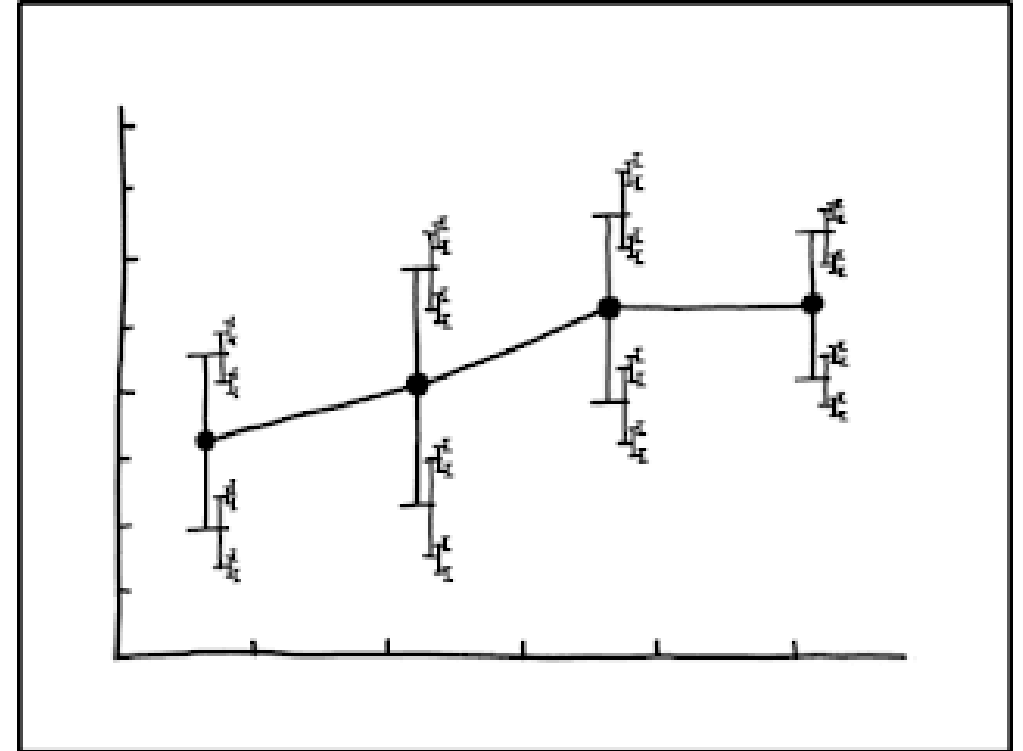
Toss The Third



make a gif.com

Primera: Propagación y cuantificación de la incertidumbre

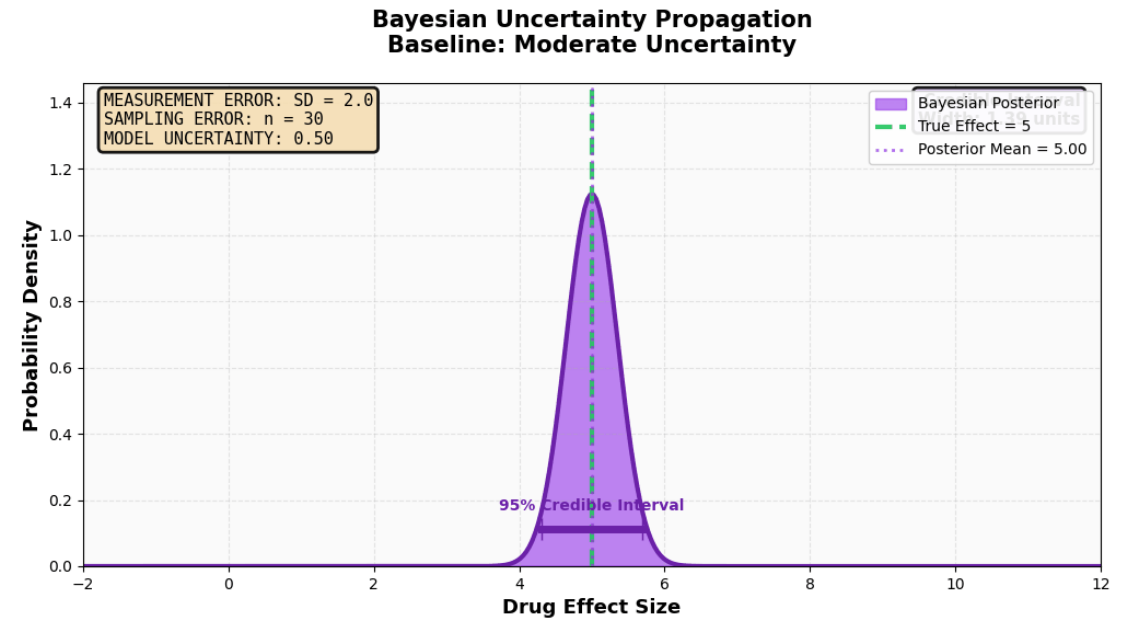
- Errores de medición
- Errores de muestreo
- Errores de especificación del modelo



I DON'T KNOW HOW TO PROPAGATE
ERROR CORRECTLY, SO I JUST PUT
ERROR BARS ON ALL MY ERROR BARS.

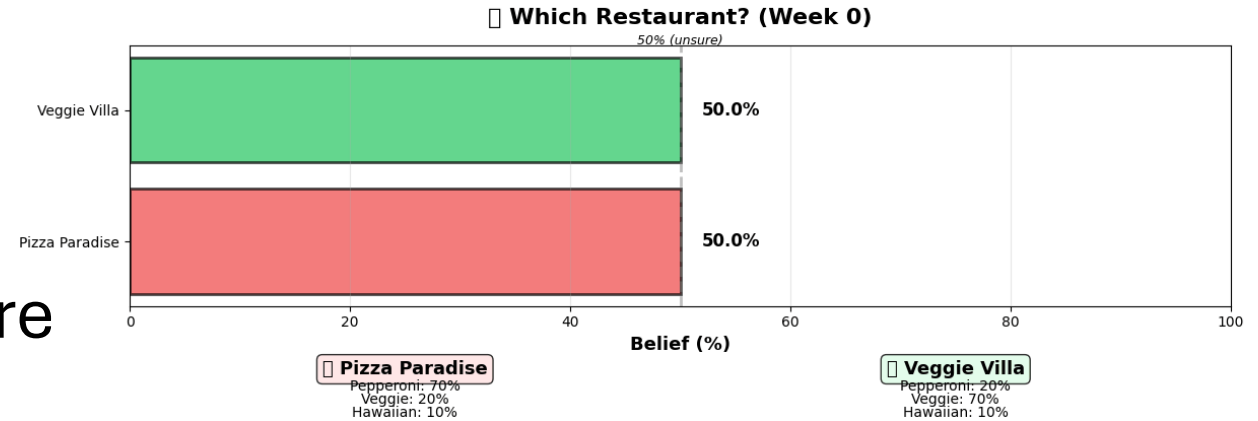
Primera

- Errores de medición
- Errores de muestreo
- Errores de especificación del modelo



Segunda: Actualización mediante la combinación de información

- Uso de *a priori*: distribuciones que contienen información sobre los parámetros de interés



Pizza History:

Pepperoni Veggie Hawaiian

Starting with 50-50 belief (no evidence yet)

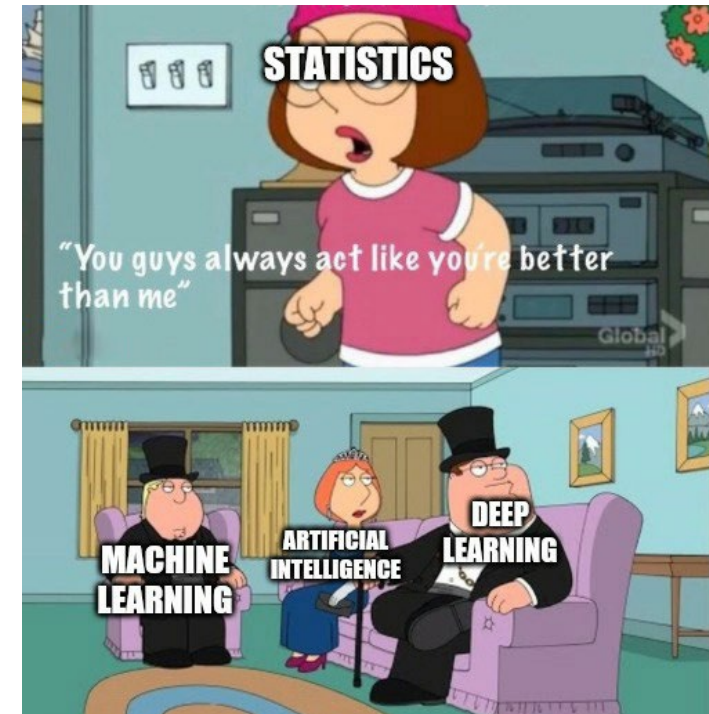


Segunda

Actualización mediante la combinación de información

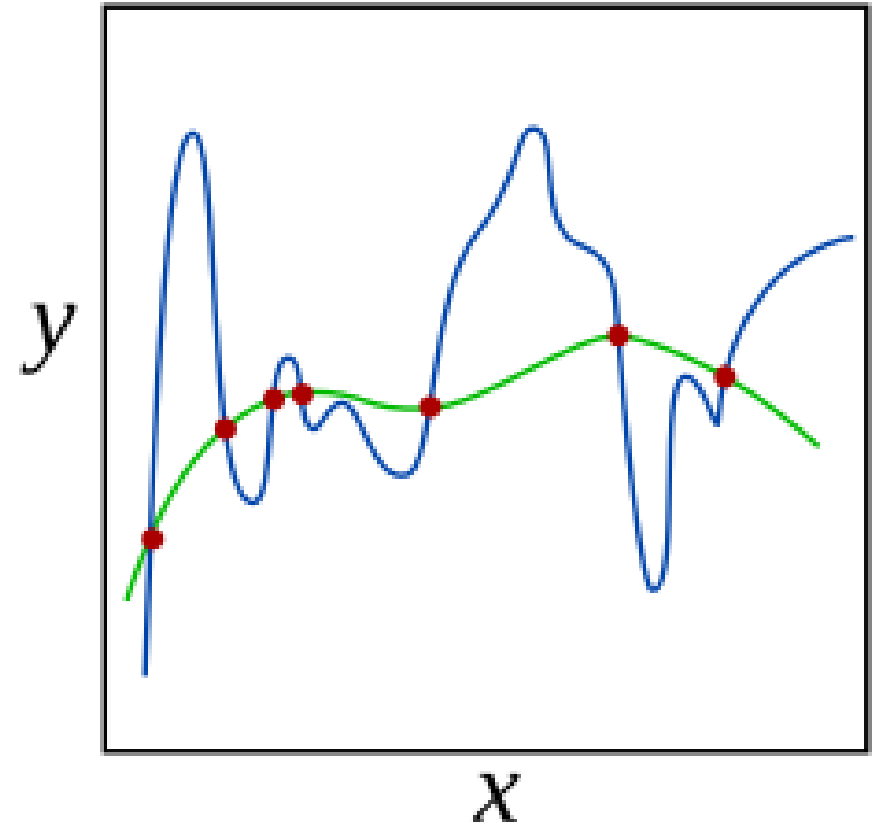
- Uso de *a priori*: distribuciones que contienen información sobre los parámetros de interés

$P(\text{next word} \mid \text{context})$

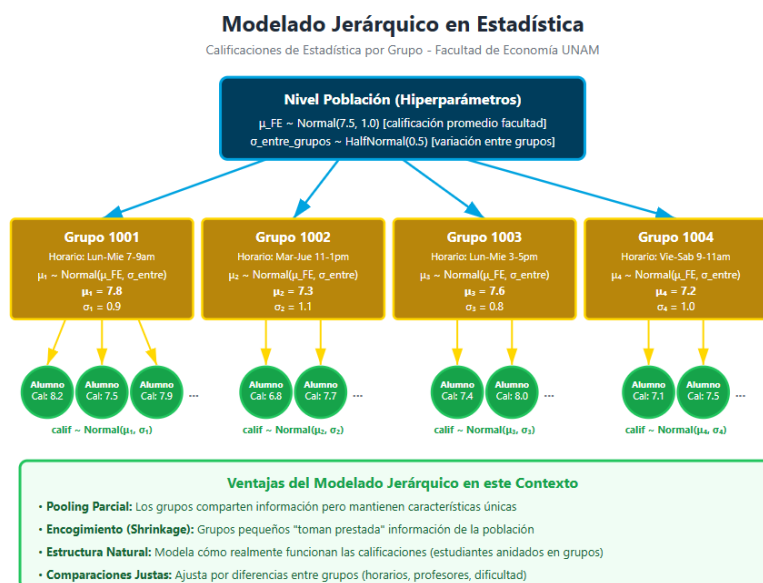


Tercera: Regularización de estimaciones

- Problemas de grandes datos, modelos complicados, y datos dispersos
- Estimaciones conservadoras
- Inferencia bayesiana permite prevenir inferencias ridículas y sobreajuste



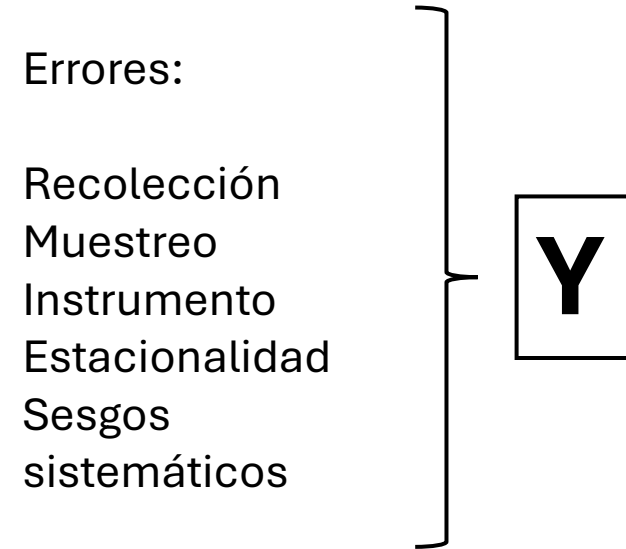
Cuarta: estructuras de datos vía parámetros



Modelar explícitamente la estructura (compleja) de las relaciones de interés

Quinta: Tratamiento de variables latentes como parámetros

- Difícilmente los datos son medidas propias de eso que nos interesa
- Usar distribuciones de parámetros para dar tratamiento a variables latentes

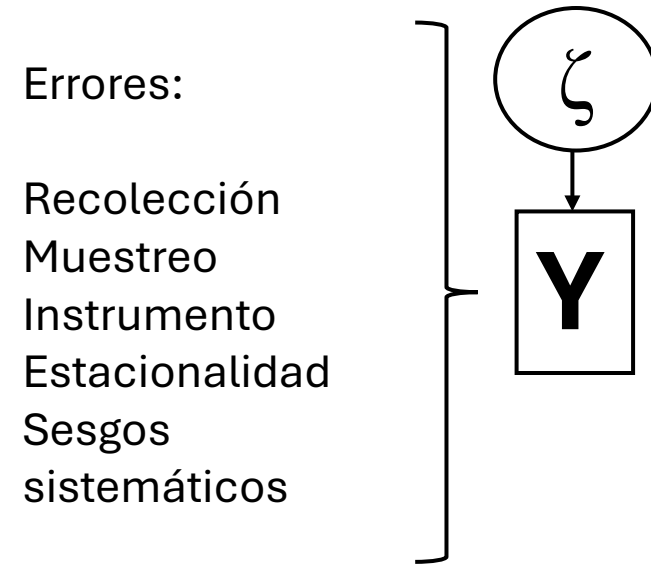


$$E = \alpha + \beta Y + \sigma$$

Quinta

Tratamiento de variables latentes como parámetros

- Considerar a los parámetros como datos latentes



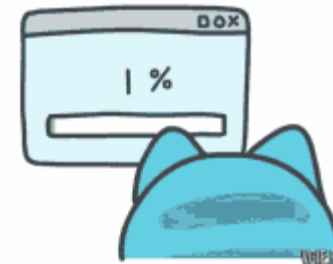
$$E = \alpha + \beta\zeta + \sigma$$

Sexta: Ir más lejos



Ir más lejos: modelos generales

- Imposibilidad de maximización (pocos-muchos datos, modelos complicados)
- Inspección de supuestos que nunca puedo evaluar directamente o cumplir en mi caso de estudio



Séptima: Inferencias directas bajo un mismo marco probabilístico

Poder concluir que:

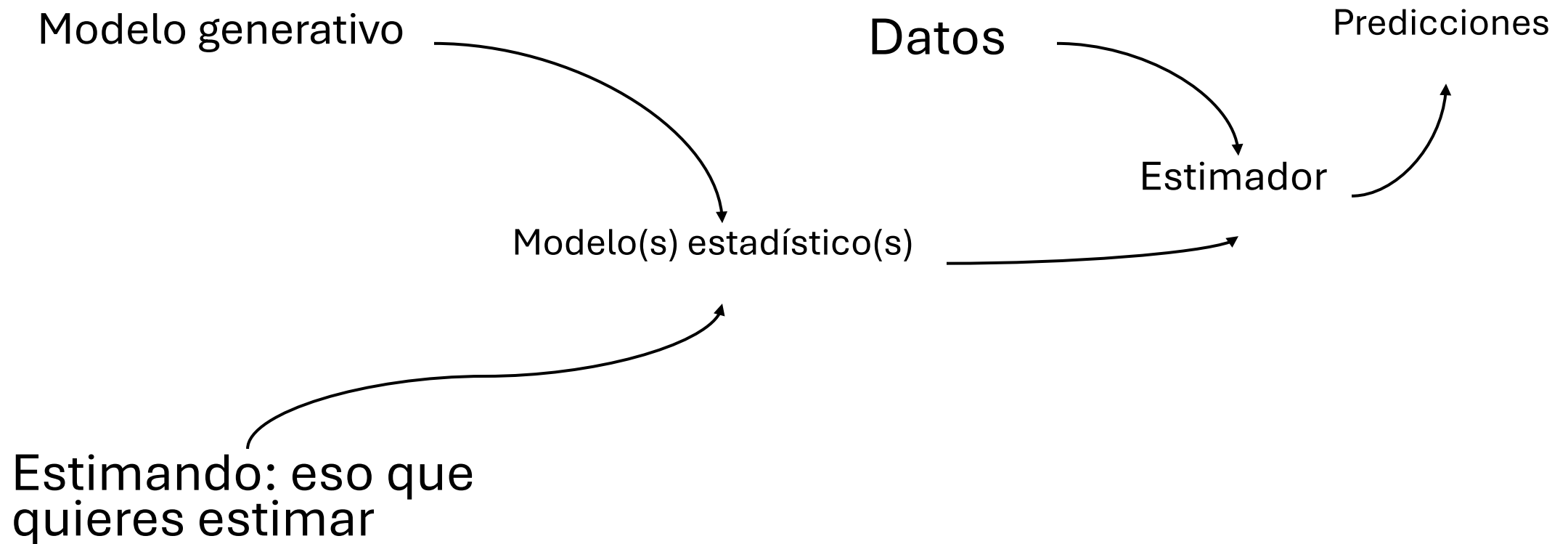
- El efecto tiene un rango probable de valores
- Que no todos los valores son igualmente probables
- Que, condicional en distintos tipos de error, el efecto probablemente es de cierto tamaño y signo

▢ INFERENCIAS BAYESIANAS
EN ECONOMÍA ▢

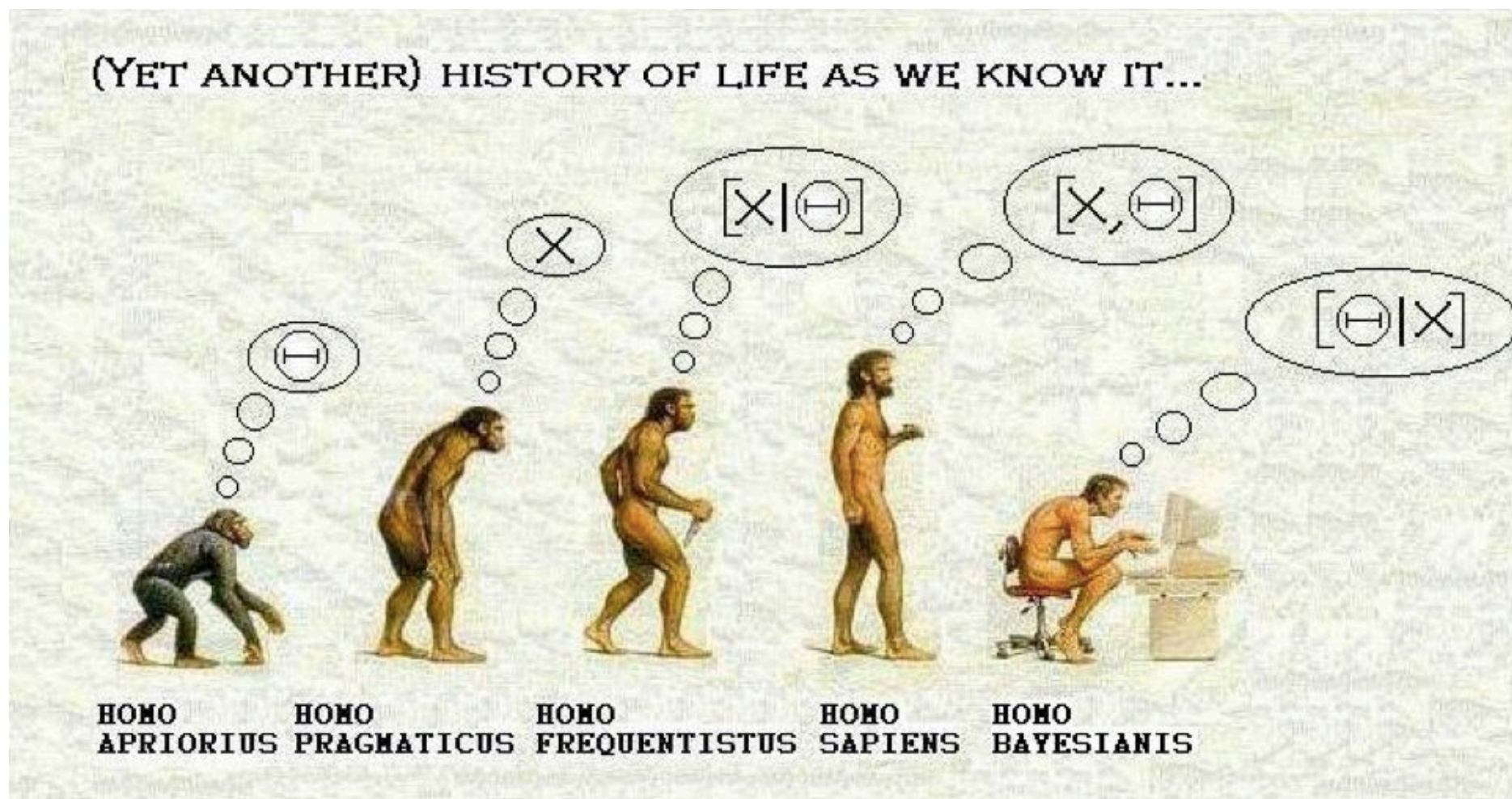
Tres casos prácticos usando
la distribución posterior completa

CASO 1: Inflación - ¿Superaremos la meta?
CASO 2: Inversión - ¿Qué proyecto elegir?
CASO 3: PIB - ¿Riesgo de recesión?

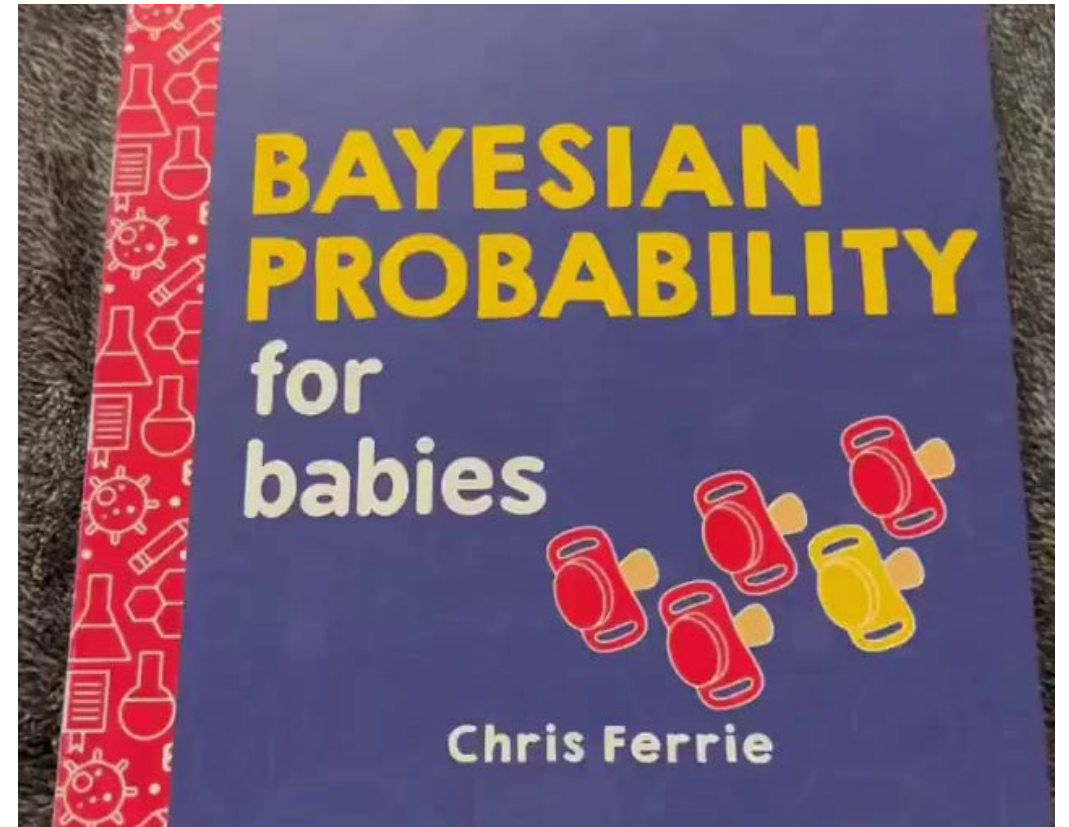
Octava: reproducibilidad, comparabilidad, diagnóstico para fallar rápido



Análisis bayesiano



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



¿Qué es la estadística bayesiana?

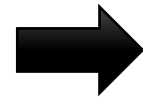
- La estadística bayesiana usa las reglas de la probabilidad para combinar toda la información disponible (datos/muestra, conocimiento previo, incertidumbre) para producir inferencias, con incertidumbre explícita, respecto al caso del uso de información por separado
- Puedes reproducir los métodos clásicos 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.

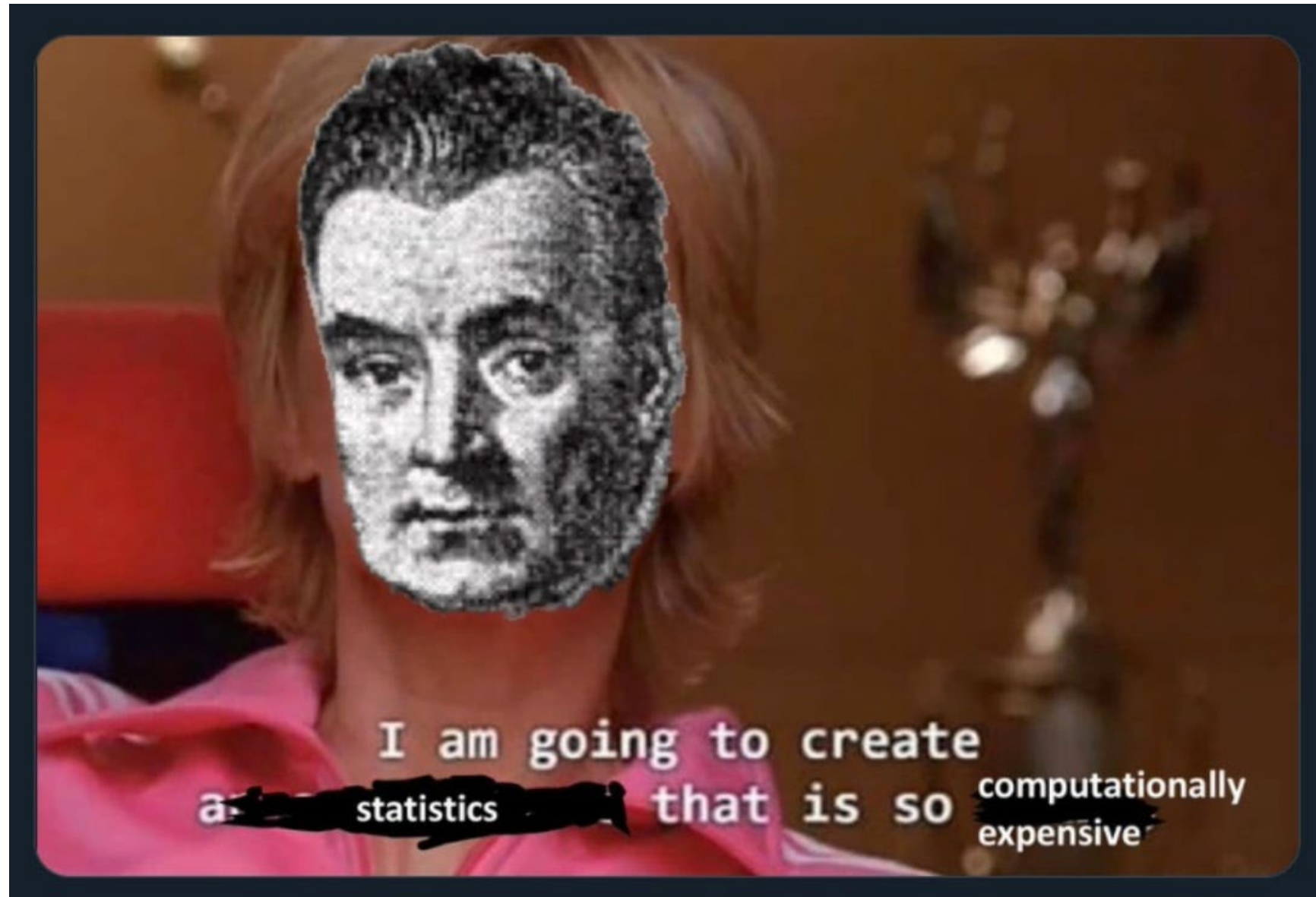
Laplace, P. S. (2012). *A philosophical essay on probabilities*. Courier Corporation.



¿De dónde viene?

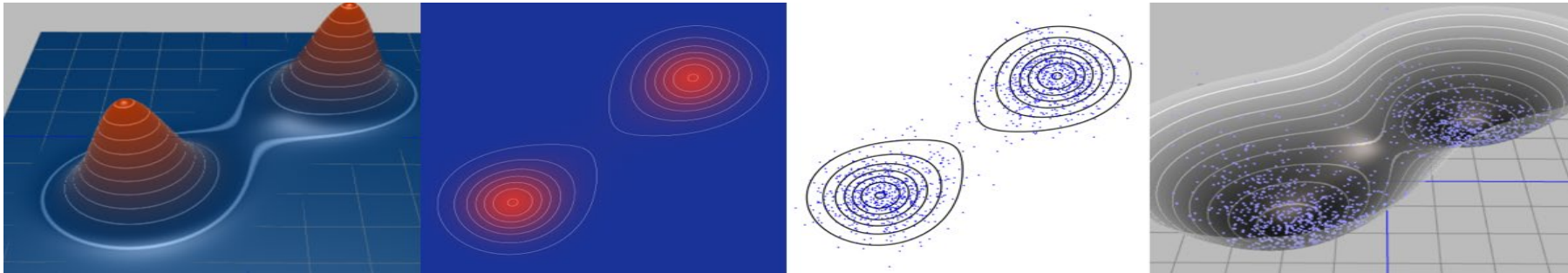
- Los métodos Bayesianos fueron usados tiempo después para resolver problemas específicos en ciencias, pero fue hasta la mitad del siglo XX que se propusieron como un marco general estadístico
- Algunas figuras clave fueron John Maynard Keynes y Frank Ramsey, quienes en los 1920s desarrollaron la teoría 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



Cómputo

- 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



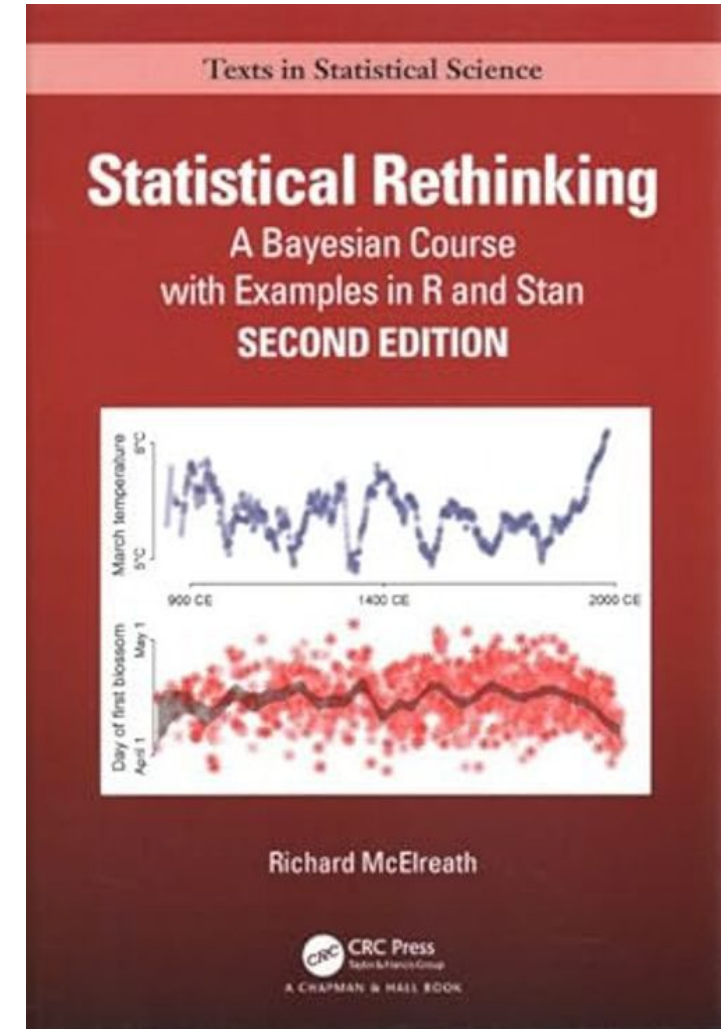
Ciencia antes que estadística



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

Estadística bayesiana

- No es que no existan otras vías para tratar *algunos* de los problemas mencionados
- Dejar de pensar a la estadística como un mero tema de la implementación de un modelo
- La estadística utiliza un conjunto de reglas y técnicas para dar tratamiento explícito a las fuentes de incertidumbre asociada a los datos, modelos, decisiones e información respecto reporte del cálculo de uno o más parámetros de interés
- La inferencia bayesiana es la única que lo hace de manera unificada



Objetivo del curso

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

Objetivos

El objetivo del curso es 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

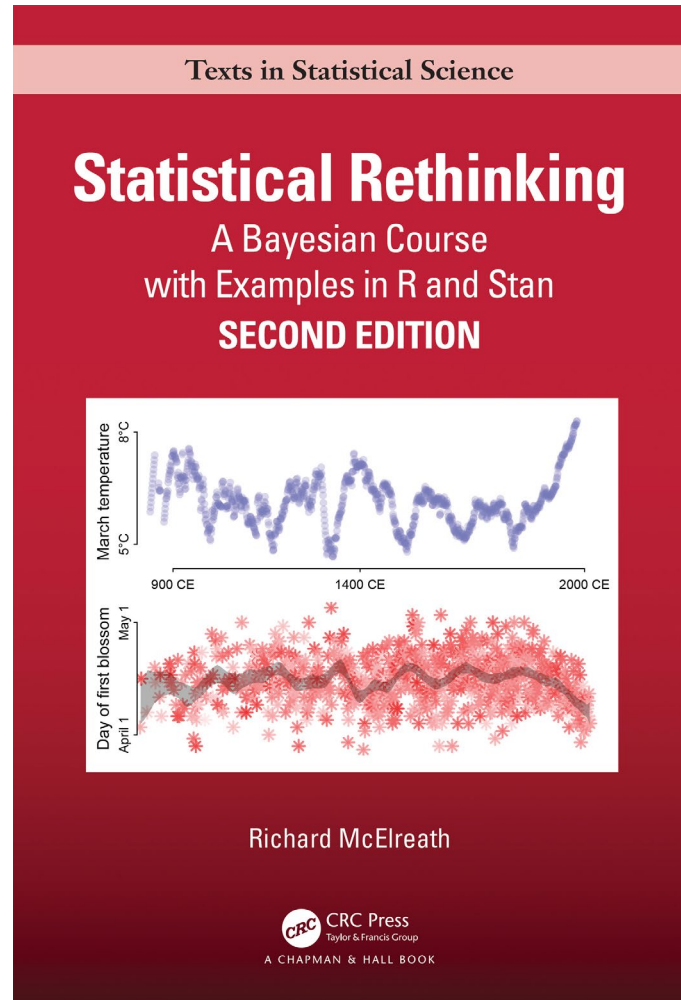
P(C): PROBABILITY THAT YOU'RE USING
BAYESIAN STATISTICS CORRECTLY

Características de las sesiones

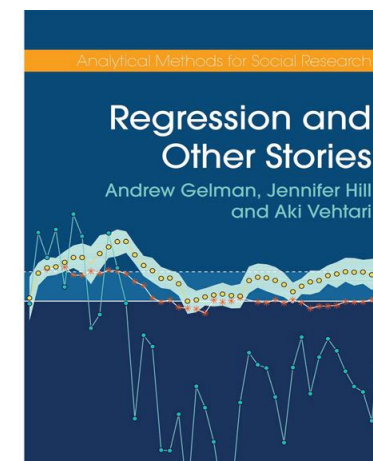
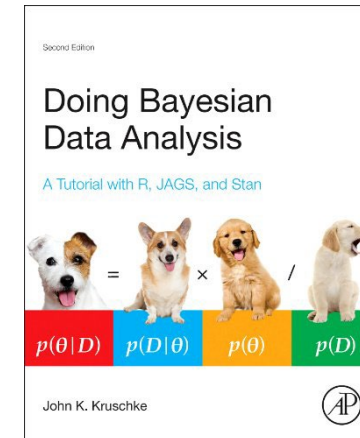
- Las sesiones combinan discusión, teoría y aplicación con el programa R.
- Antes de cada clase, los alumnos deberán leer una selección de capítulos/artículos para su discusión en grupo.
- Los docentes impartirán cada sesión (prepararán un archivo.ppt que subirán a Github después de cada clase) y se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

Bibliografía

Básica



Secundaria



Evaluación

Se utilizará un ejercicio 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*)

Temario (General)

2. Fundamentos aplicados a la inferencia de probabilidades

Inferir probabilidades binomiales; técnicas de Monte Carlo vía cadenas de Markov; modelos jerárquicos; abordajes bayesianos a la prueba de hipótesis

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/ClaseBayes2026/>

Lugar y hora

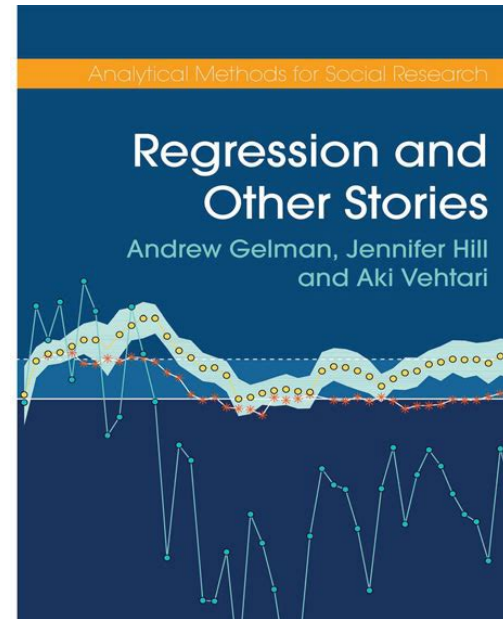
16 sesiones los jueves de 16:00 a 19:00

Próxima clase

Esencial:

Capítulo 1. Regression and other stories.

Gelman, Hill y Vehtari.



Próxima clase

- 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 revolutionized

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 Pres.

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!*

