



UNAM  
POSGRADO



Programa  
Universitario  
de Estudios  
del Desarrollo  
UNAM

# Inferencia estadística

## Modelaje bajo incertidumbre

Dr. Héctor Nájera  
Dr. Curtis Huffman



Usar modelos matemáticos para hacer afirmaciones generales a partir de un conjunto particular de datos

## **INFERENCIA ESTADÍSTICA**

# ¿Las siete tramas básicas?

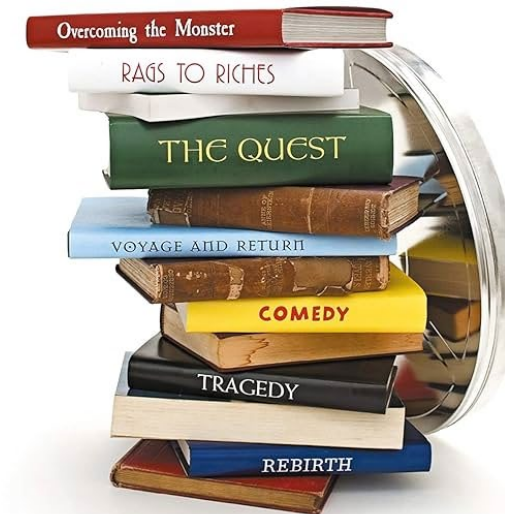
Booker, C. (2004). *The seven basic plots: Why we tell stories*. A&C Black.

- Superando al monstruo
- Pobreza a la riqueza
- La búsqueda
- Viaje y regreso
- Comedia
- Tragedia
- Renacimiento

'This is the most extraordinary, exhilarating book.' FAY WELDON

## THE SEVEN BASIC PLOTS

*Why we tell stories*



CHRISTOPHER  
BOOKER

BLOOMSBURY



# Los 3 retos de la inferencia estadística

- Generalizar de la muestra a la población (muestreo: la muestra no es la población)
- Generalizar del grupo de tratamiento al control (causalidad: el tratamiento no es el control)
- Generalizar de las indicaciones instrumentales al concepto de interés (medición: las indicaciones instrumentales no son los resultados de medición)

Los tres retos pueden enmarcarse como problemas de **predicción**

- Nuevas personas o ítems que no están en la muestra
- Resultados de una posible intervención
- Objetos científicos si éstos fuesen observable



# ¿Dijo usted incertidumbre?

Tres paradigmas diferentes

- Errores de muestreo (modelo muestral)
- Errores de medición (modelo de medición)
- Errores de modelaje (relaciones supuestas entre los datos)

Noten que

- Se puede incurrir en errores de muestreo con mediciones “perfectas”
- Se puede incurrir en errores de medición aun cuando se trabaja con la población entera
- Se puede incurrir en errores de modelaje (no sabemos cuál modelo usar) aun con mediciones perfectas y la población entera





## 3 preocupaciones del modelaje matemático

- ¿Qué información es utilizada en el proceso de estimación?
  - Conozca sus datos (siempre incompletos e imperfectos)
- ¿Qué supuestos son necesarios?
  - Conozca su modelo (¿predice sinsentidos?)
- ¿Cómo se interpretan las estimaciones y sus predicciones?
  - ¿Qué permite afirmar lo modelado?

# Modelaje probabilístico

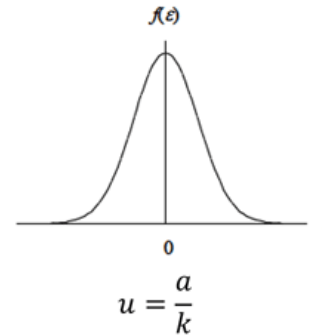
- Las distribuciones representan los aspectos no modelados del sistema bajo medición (incertidumbre)
- “[Randomness] is, evidently, a euphemism, whose real meaning is: deliberately throwing away relevant information when it becomes too complicated for us to handle.” Edwin T. Jaynes



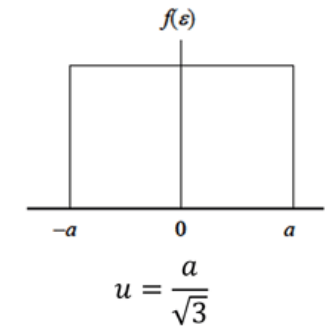
# Distribuciones de probabilidad

- Representación matemática de
  - **información** disponible
  - verosimilitud (probabilidad) de los diferentes valores
  - Proceso generador de datos
- Componente fundamental de la inferencia estadística
  - Modelo/algorithm predictivo

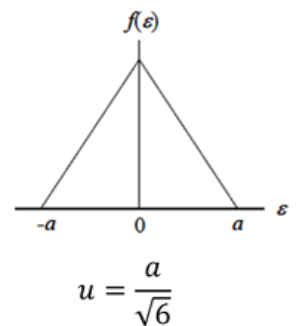
Normal



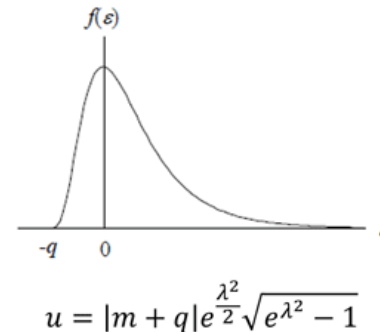
Uniform



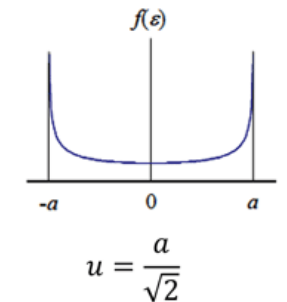
Triangle



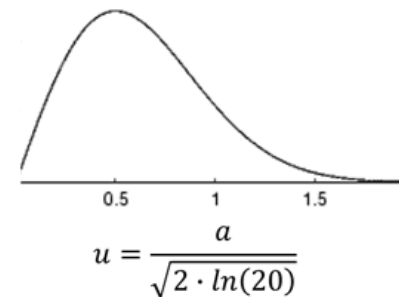
Log Normal



U-Shaped



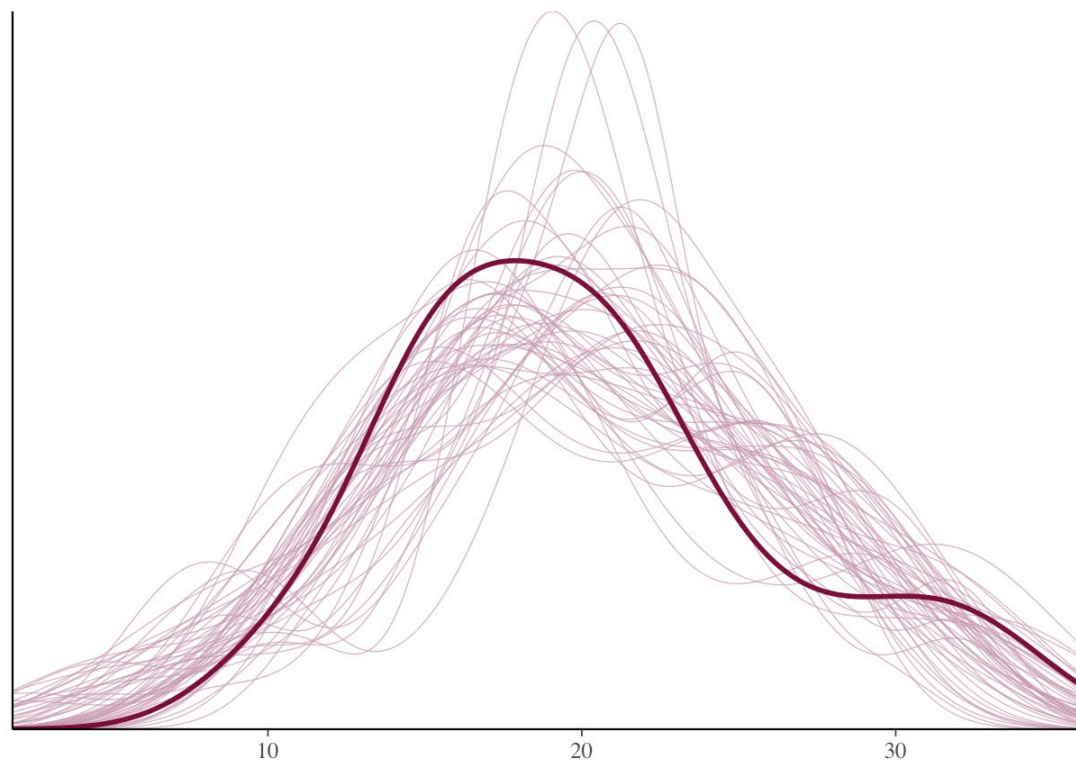
Rayleigh



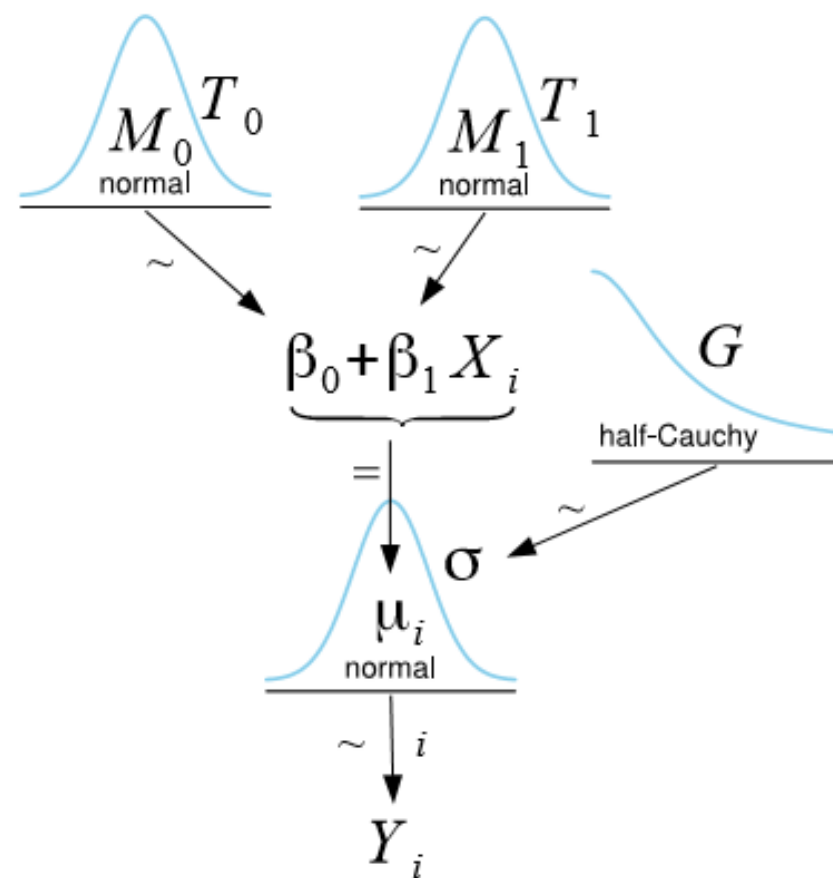


# Modelos generativos (primordiales)

- Usamos distribuciones **lo mismo** para representar **datos** que sus **parámetros** en un modelo generativo de todas las variables.
- Distribuciones de parámetros de distribuciones de parámetros de...



—  $y$   
—  $y_{\text{rep}}$



- El análisis bayesiano involucra describir datos por medio de modelos matemáticos, y asignar credibilidad a valores de parámetros que son consistentes con los datos y el conocimiento previo.
- El abordaje bayesiano formaliza la reubicación de credibilidad sobre posibilidades ante nuevos datos.

Formula:

$$f(\text{model} \mid \text{data}) = \frac{f(\text{data} \mid \text{model}) \times f(\text{model})}{f(\text{data})}$$

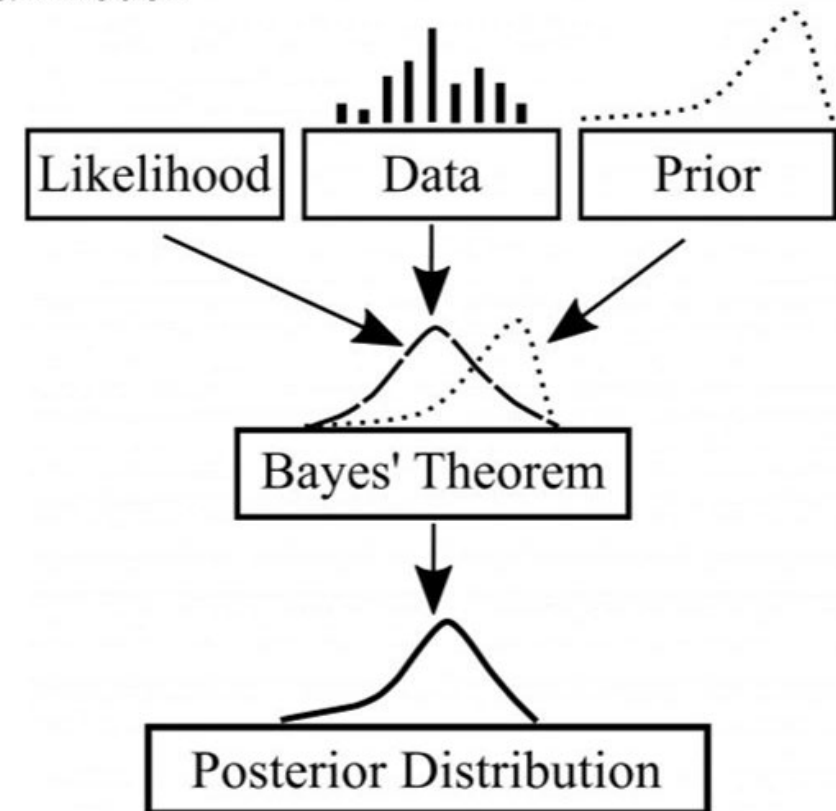
Bayesian Model

Likelihood distribution

Prior distribution

Data

Bayesian Model:



# Sociología de la economía

- Los métodos tradicionales (frecuentistas) que usan sólo distribuciones muestrales son prácticos y útiles en muchos problemas idealizados y particularmente simples
  - Los datos tienen distribuciones, los parámetros no
  - La *likelihood* no es una distribución de probabilidad
  - (Super) Poblaciones imaginarias para eventos no repetibles

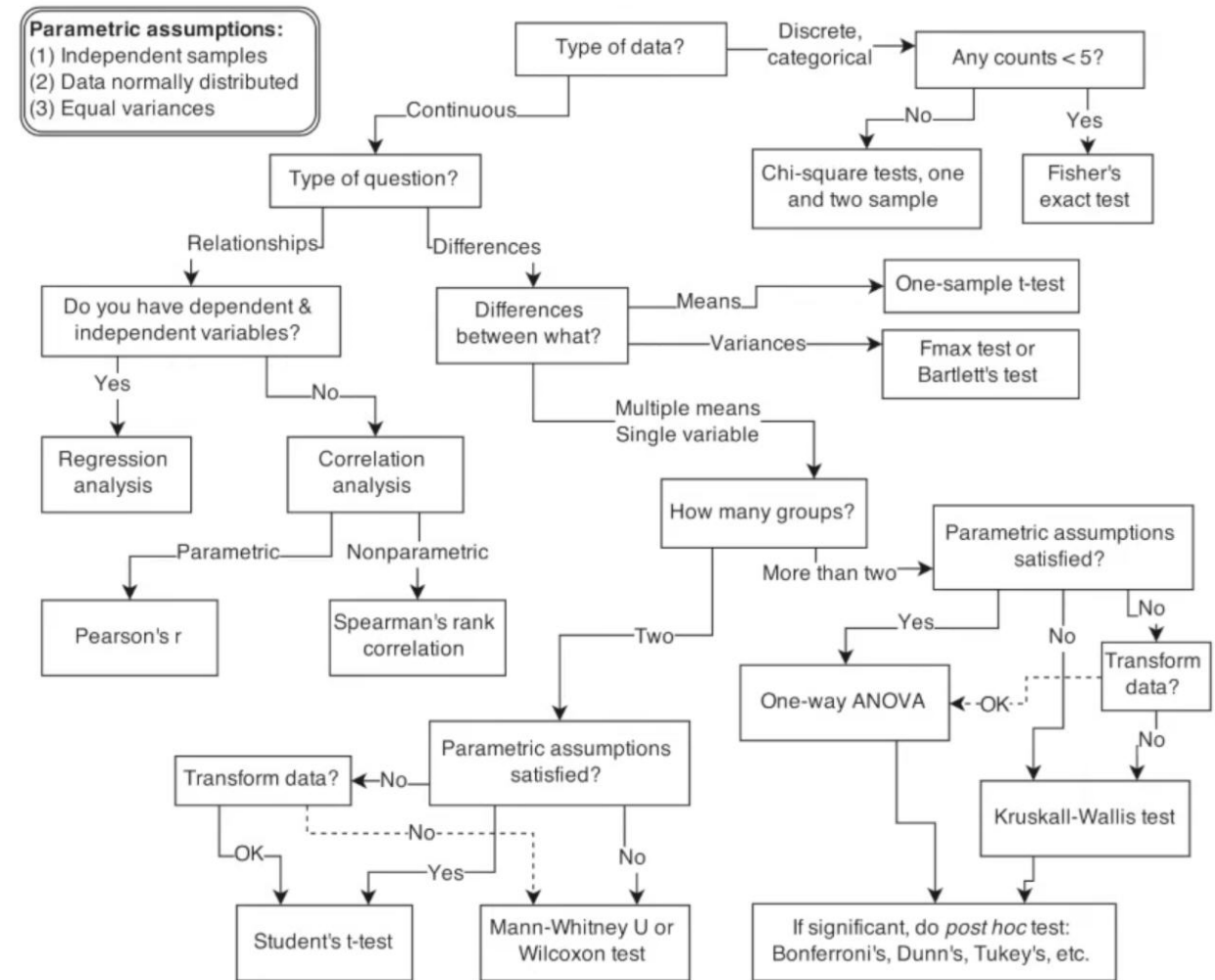


- Presuponen condiciones (repeticiones independientes sin información previa relevante) que raramente se observan en problemas de investigación.
- Este abordaje se ha probado inadecuado ante las necesidades actuales de la investigación empírica





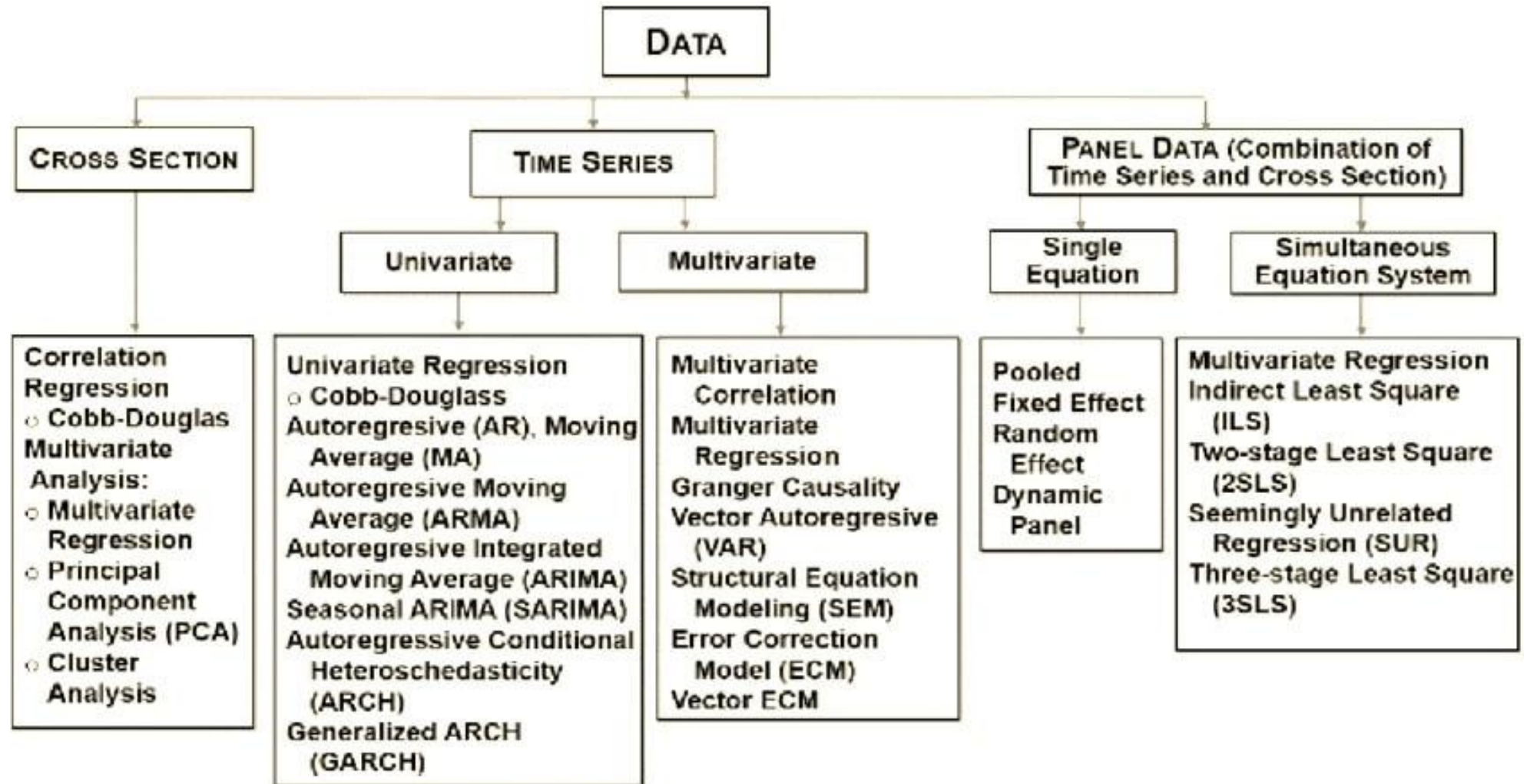
- Los métodos frecuentistas obligar ajustar la investigación al estadíst escogido antes que a modelar el problema en cuestión, para luego inventar artefactos *ad hoc* intervalos de confianza y pruebas significancia.
- La mayor parte (si no es que todo de su entrenamiento estadístico h sido en el paradigma frecuentista.





	<i>Hypothesis</i>	<i>Statistical Tools Applied</i>
1	The Public organizations are under-using IT potentials in comparison to private sector.	t-test
2	Escalation (time-delays, cost-overruns, compromise on lesser objectives) of IT projects is more common in public organizations than in private enterprises.	t-test
3	IT-people overestimate while non-IT workers underestimate the role of IT in the organizations.	t-test
4	Public sector is less optimistic about the role of IT than private sector.	t-test
5	Professors, doctors and consultants view IT differently.	ANOVA
6	Experience of non-IT workforce is negatively correlated with perceptions about IT.	Correlation analysis
7	Higher the perceptions about IT, greater are the chances/perceptions of success in IT projects	Simple Regression
8	The organizational, human, contextual and technological factors collectively determine the variation in the success/failure of an IT-project.	Multiple Regression







# Diagrama frecuentista de construcción de conocimiento

---





**Justin Stewart**  
@thecrobe

the p in p-value stands for please use a bayesian approach instead

8:57 AM · Aug 28, 2020 · Twitter for iPhone

<https://twitter.com/thecrobe/status/1299330167920373761>



# Construya, estime, critique, repita

- El mundo se transforma en datos sólo de manera imperfecta. Si hemos de aprender algo de los datos sobre el mundo, necesitamos buscar y entender sus imperfecciones y las implicaciones de esas imperfecciones.
  - Unidad entre variables: sin distinciones entre datos y parámetros anteriores a la observación.
  - Unidad entre distribuciones: sin distinciones entre likelihoods y priors anteriores la observación.
  - Modelo generativo anterior en sentido lógico a la observación
    - Complica el flujo de trabajo
    - Provee estructura
    - La investigación empírica ES el desarrollo de modelos estadísticos
    - Tenemos que averiguar cómo enseñar el flujo de trabajo (se está investigando)





# Próxima sesión

Esencial:

Kruschke, J. K. (2013). Bayesian estimation supersedes the  $t$  test. *Journal of Experimental Psychology: General* 142(2), 573.

Journal of Experimental Psychology: General  
2013, Vol. 142, No. 2, 573–603

© 2012 American Psychological Association  
0096-3445/13/\$12.00 DOI: 10.1037/a0029146

## Bayesian Estimation Supersedes the $t$ Test

John K. Kruschke  
Indiana University, Bloomington

Bayesian estimation for 2 groups provides complete distributions of credible values for the effect size, group means and their difference, standard deviations and their difference, and the normality of the data. The method handles outliers. The decision rule can accept the null value (unlike traditional  $t$  tests) when certainty in the estimate is high (unlike Bayesian model comparison using Bayes factors). The method also yields precise estimates of statistical power for various research goals. The software and programs are free and run on Macintosh, Windows, and Linux platforms.

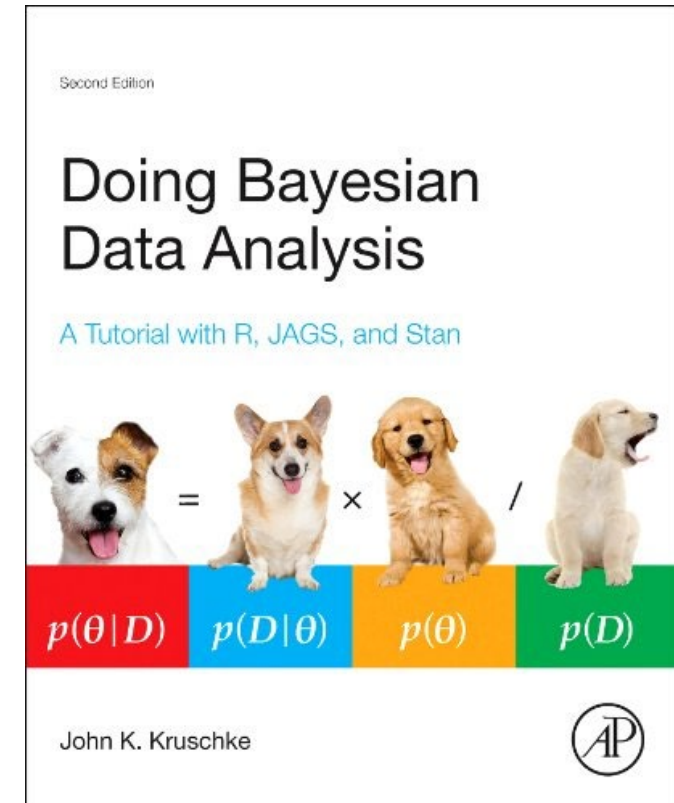
*Keywords:* Bayesian statistics, effect size, robust estimation, Bayes factor, confidence interval

# Próxima sesión

Recomendada:

Kruschke, J. (2014). Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan.

2. Introduction: Credibility, Models, and Parameters
4. What is This Stuff Called Probability?





# CONTACTO

Dr. Héctor Nájera y Dr. 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](mailto:hecatalan@hotmail.com), [chuffman@unam.mx](mailto:chuffman@unam.mx)



**¡Bienvenido  
s  
estudiantes  
!**

