

MGE-201: Inferencia estadística

Felipe Osorio

felipe.osorio@uv.cl

Información

Horario:

Clases: Jueves, 18:00-21:00 hrs. via TEAMS¹.

Contacto:

E-mail: felipe.osorio@uv.cl.

Web: <https://github.com/faosorios/Curso-Inferencia> y **AULA**

Evaluación:

Se realizará **2 Pruebas**: 9-Oct, 4-Dic, **Tareas** y un **Reporte final**.

¹Coordenadas en AULA

Criterio de aprobación

Criterio de aprobación:

La nota final (NF) es calculada como sigue:

$$NF = 0.3 \cdot P_1 + 0.3 \cdot P_2 + 0.2 \cdot \bar{T} + 0.2 \cdot R,$$

donde P_1 y P_2 representan las notas en las pruebas 1 y 2, mientras que \bar{T} representa el promedio de tareas y R corresponde a la nota del reporte final.

Aquellos estudiantes que obtengan NF mayor o igual a 40, aprobarán la asignatura.

Reglas adicionales

- ▶ Se puede realizar [preguntas](#) sobre la materia en [cualquier momento](#).
- ▶ En formato online, los alumnos deben [apagar/silenciar](#) sus [micrófono](#) durante clases.
- ▶ Conversaciones sobre asuntos ajenos a la clase no serán tolerados. Otros estudiantes tiene derecho a [asistir clases en silencio](#).
- ▶ Al enviar algún [e-mail](#) al profesor, identificar el código de la asignatura en el asunto ([MGE201](#)).
- ▶ [E-mail](#) será el canal de [comunicación oficial](#) entre el profesor y los estudiantes.

Reglas: sobre las evaluaciones

- ▶ Es derecho del estudiante conocer la pauta de corrección la que será publicada en la página web del curso.
- ▶ Use principalmente lapiz pasta (no utilice lapiz rojo).
- ▶ Pedidos de recorrección deben ser argumentados por escrito.
- ▶ En modalidad online, Certámenes deben ser enviados en formato PDF.²
- ▶ Cualquier tipo de fraude en prueba (copia, uso de WhatsApp, suplantación, etc.) será sancionado.

²En un único archivo, orientado en una dirección legible.

Orientaciones de estudio

- ▶ Mantener la frecuencia de estudio de inicio a final del semestre. El ideal es estudiar el contenido luego de **cada** clase.
- ▶ Estudiar primeramente el contenido dado en clases, **buscando apoyo en las referencias bibliográficas**.
- ▶ Las **referencias son fuentes de ejemplos y ejercicios**. Resuelva una buena cantidad de ejercicios. **No deje esto para la víspera de la prueba**.
- ▶ Buscar las **referencias bibliográficas** al inicio del semestre, dando preferencia a las principales y complementarias.

Programa del curso³

- ▶ Introducción y conceptos básicos.
- ▶ Familia exponencial y sus propiedades.
- ▶ Métodos de estimación.
- ▶ Test de hipótesis: test de Neyman-Pearson y test asintóticos.
- ▶ Inferencia Bayesiana.

³Este es un curso **fundamental** donde exploramos métodos para abordar el proceso de inferencia, **no** es un curso enfocado **exclusivamente** en el análisis de datos.

Bibliografía

-  Azzalini, A. (1996).
Statistical Inference Based on the Likelihood.
Chapman & Hall, London.
-  Casella, G., and Berger, R.L. (2002).
Statistical Inference, 2nd Ed..
Duxbury, Pacific Grove.
-  Mood, A.M., Graybill, F.A., Boes, D.C. (1974).
Introduction to the Theory of Statistics.
McGraw-Hill, New York.
-  Robert, C.P. (2007).
The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation, 2nd Ed.
Springer, New York.

Ideas subyacentes

“Todos los modelos son errados, pero algunos son útiles.”

– George Box.

“Aunque puede parecer una paradoja, toda la ciencia exacta está dominada por la idea de aproximación.”

– Bertrand Russell.

Principio KISS: “Keep It Simple, Stupid.”

– Clarence “Kelly” Johnson.

Ideas subyacentes

“Todos los modelos son errados, pero algunos son útiles.”

– George Box.

“Aunque puede parecer una paradoja, toda la ciencia exacta está dominada por la idea de aproximación.”

– Bertrand Russell.

Principio KISS: “Keep It Short and Simple.”

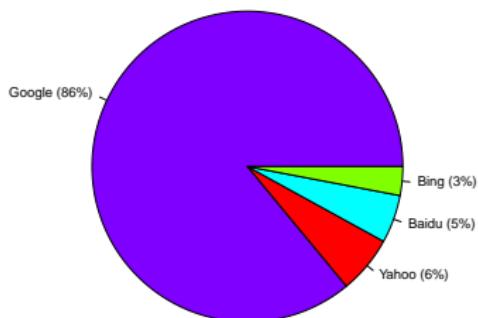
– Clarence “Kelly” Johnson.

El éxito de Google: Aplicar el principio KISS⁴

Evolución de Yahoo vs. Google:



Cuota de mercado de los motores de búsqueda:



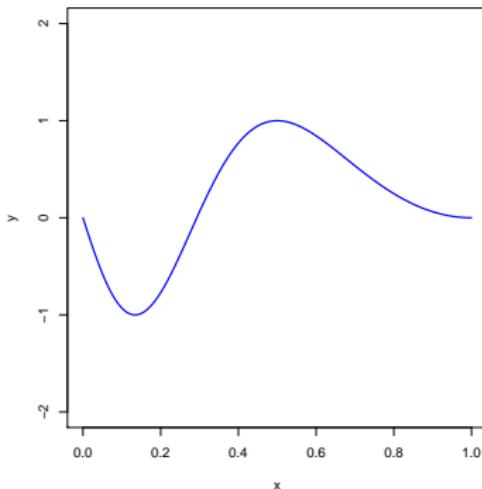
⁴En estadística este se conoce como Principio de Parsimonia.

El problema del modelado

Considere la función

$$Y = \operatorname{sen}\{2\pi(1 - x)^2\},$$

cuyo gráfico es dado por:

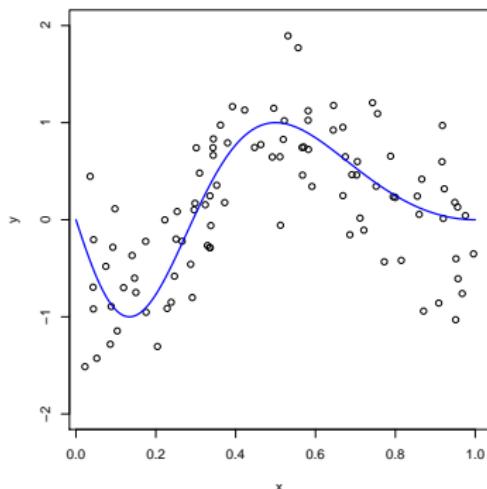


El problema del modelado

Suponga que “generamos” datos, usando

$$Y_i = \sin\{2\pi(1 - x_i)^2\} + \sigma\epsilon_i, \quad i = 1, \dots, 100,$$

donde $x_i \sim U(0, 1)$, $\epsilon_i \sim N(0, 1)$ y $\sigma = 1/2$,

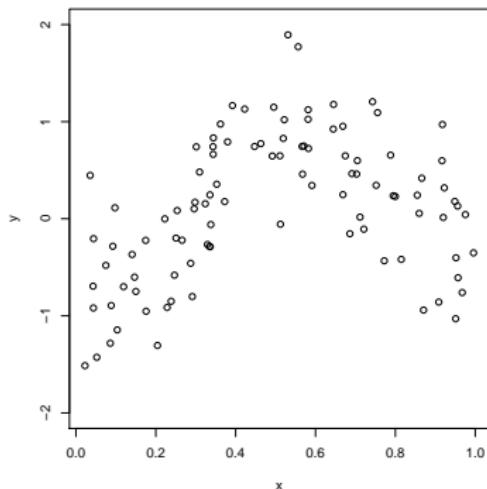


El problema del modelado

Lamentablemente, en la práctica **sólo** disponemos de los **datos observados**:

$$(x_1, Y_1), (x_2, Y_2), \dots, (x_{100}, Y_{100}),$$

el primer paso es hacer un análisis exploratorio:

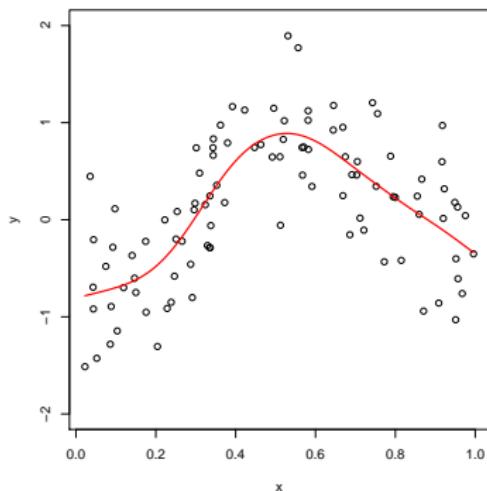


El problema del modelado

El analista propone el **modelo**:

$$Y_i = g(x_i) + \epsilon_i, \quad i = 1, \dots, 100,$$

y su objetivo es “estimar” la función $g(\cdot)$ desde los datos, obteniendo

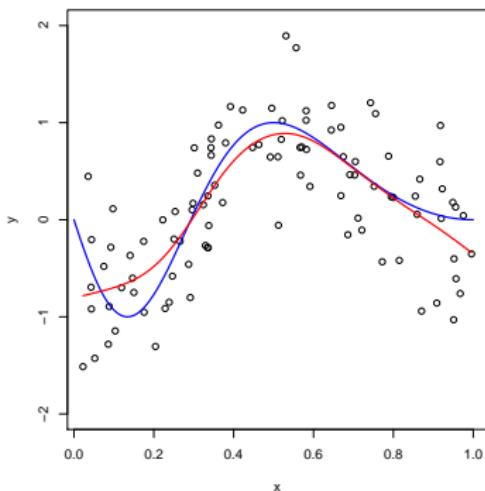


El problema del modelado

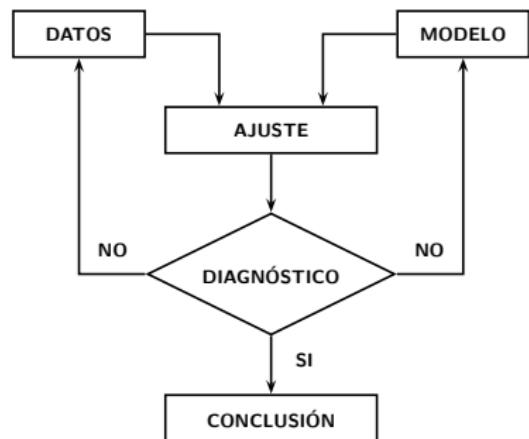
En Estadística se estudia teóricamente, la “bondad del modelo” comparando

$$\hat{Y} = \hat{g}(x), \quad \text{v.s.} \quad Y = \sin\{2\pi(1-x)^2\},$$

esto es, el **modelo ajustado** v.s. el **modelo subyacente** (verdadero).



Esquema de Modelación Estadística



Recolección de datos: **Muestreo.**

Análisis exploratorio de datos.

Análisis Multivariado.

Técnicas de Regresión.

Series de Tiempo, entre (muchas) otras.

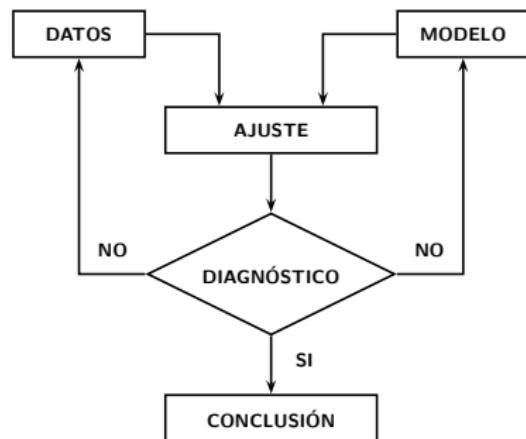
Inferencia Estadística.

Bondad de ajuste, técnicas gráficas.

Ánalisis de Sensibilidad.

Comunique sus resultados!

Esquema de Modelación Estadística



Recolección de datos: **Muestreo**.

Análisis exploratorio de datos.

Análisis Multivariado.

Técnicas de Regresión.

Series de Tiempo, entre (muchas) otras.

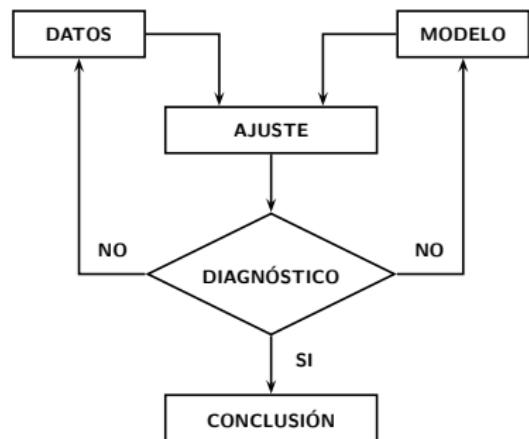
Inferencia Estadística.

Bondad de ajuste, técnicas gráficas.

Análisis de Sensibilidad.

Comunique sus resultados!

Esquema de Modelación Estadística



Recolección de datos: **Muestreo**.

Análisis exploratorio de datos.

Análisis Multivariado.

Técnicas de Regresión.

Series de Tiempo, entre (muchas) otras.

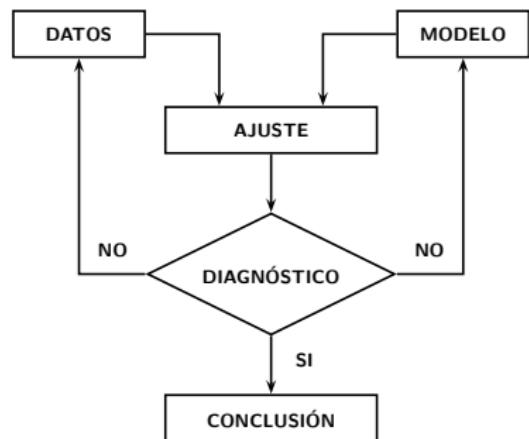
Inferencia Estadística.

Bondad de ajuste, técnicas gráficas.

Análisis de Sensibilidad.

Comunique sus resultados!

Esquema de Modelación Estadística



Recolección de datos: **Muestreo**.

Análisis exploratorio de datos.

Análisis Multivariado.

Técnicas de Regresión.

Series de Tiempo, entre (muchas) otras.

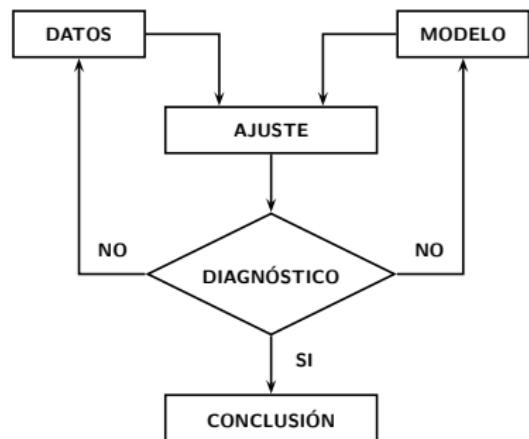
Inferencia Estadística.

Bondad de ajuste, técnicas gráficas.

Análisis de **Sensibilidad**.

Comunique sus resultados!

Esquema de Modelación Estadística



Recolección de datos: **Muestreo**.

Análisis exploratorio de datos.

Análisis Multivariado.

Técnicas de Regresión.

Series de Tiempo, entre (muchas) otras.

Inferencia Estadística.

Bondad de ajuste, técnicas gráficas.

Análisis de **Sensibilidad**.

Comunique sus resultados!