

MAT-206: Inferencia Estadística

Felipe Osorio

`fosorios.mat.utfsm.cl`

Departamento de Matemática, UTFSM



Horario:

Clases: Lunes y Miércoles, bloque 3-4 (09:35-10:45 hrs.) via Sala F-249.

Contacto:

E-mail: felipe.osorios@usm.cl.

Material de clases:

Página del curso (GitHub): <https://github.com/faosorios/Curso-Inferencia>

Página personal: <http://fosorios.mat.utfsm.cl/teaching.html#MAT206>

Evaluación:

Se realizará **3 Certámenes**.¹

¹28-Sep, 09-Nov, 12-Dic.



Criterio de aprobación:

Sea NP el promedio de los Certámenes. Aquellos estudiantes que obtengan NP mayor o igual a 55 y todos los certámenes sobre 40, aprobarán la asignatura con nota final, $NF = NP$.

Criterio para rendir global:

En caso contrario, y siempre que $NP \geq 45$,² los estudiantes podrán rendir el certamen global (CG), en cuyo caso la nota final es calculada como sigue:

$$NF = 0.6 \cdot NP + 0.4 \cdot CG.$$

²Si $NP < 45$ usted ha reprobado la asignatura.



Reglas adicionales

- ▶ Se llevará un **control de asistencia**.
- ▶ Se puede realizar **preguntas** sobre la materia en **cualquier momento**.
- ▶ 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**.
- ▶ Alumnos que lleguen tarde o se retiren deben hacerlo en **silencio**.
- ▶ Al enviar algún **e-mail al profesor**, identificar el código de la asignatura en el asunto (**MAT206**).
- ▶ **E-mail** será el canal de **comunicación oficial** entre el profesor y los estudiantes.



Reglas: sobre los certámenes

- ▶ Todas las **hojas necesarias** para responder las pruebas **serán entregadas por el profesor**.
- ▶ Será permitido el uso de una **calculadora científica simple** (no del celular).
- ▶ 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 corrección **deben ser argumentados por escrito**.
- ▶ **Cualquier tipo de fraude en prueba** (copia, uso de WhatsApp, suplantación, etc.) será llevado a **Comisión Universitaria**.



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.



Prerrequisitos





- ▶ El requisito formal es [MAT-263: Teoría de Probabilidades y Procesos Estocásticos](#).
- ▶ En este curso se profundiza y extiende los contenidos revisados en [MAT-041: Probabilidad y Estadística](#).
- ▶ Se asume un conocimiento básico de los siguientes aspectos:
 - ▶ Variables aleatorias.
 - ▶ Convergencia de variables aleatorias.
 - ▶ Manipulación de matrices y vectores aleatorios (usaremos algunos resultados matriciales básicos).



- ▶ Introducción y conceptos básicos.
- ▶ Familia exponencial y sus propiedades.
- ▶ Métodos basados en la verosimilitud.
- ▶ Métodos basados en modelos parcialmente especificados.
(MM, M -estimadores, extremum estimators, funciones de inferencia)
- ▶ Regiones de confianza.
- ▶ Test de hipótesis: test de Neyman-Pearson y test asintóticos.

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



-  Casella, G., and Berger, R.L. (2002).
Statistical Inference (2nd Ed.).
Duxbury, Pacific Grove.
-  Jørgensen, B., and Labouriau, R. (1994).
Exponential Families and Theoretical Inference.
Lecture Notes, Department of Statistics, University of British Columbia.
-  Pawitan, Y. (2001).
In All Likelihood: Statistical Modelling and Inference using Likelihood.
Oxford University Press.
-  Rohde, C.A. (2014).
Introductory Statistical Inference with the Likelihood Function.
Springer, New York.

- ▶ Página de la asignatura: <https://github.com/faosorios/Curso-Inferencia>
- ▶ Notas de clases: [Elementos de Inferencia Estadística \(PDF, 65 pág.\)](#)
- ▶ Material de clases: Archivos PDF con las principales temáticas abordadas.
- ▶ Evaluaciones de años anteriores (2017, 2018, 2020, 2021) con sus respectivas pautas de corrección.



"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.

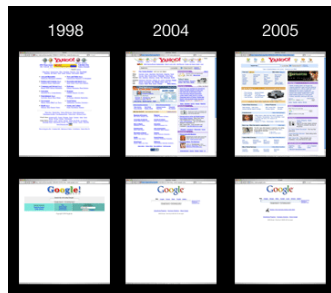
Principio KISS: "Keep It Short and Simple."

– Clarence "Kelly" Johnson.

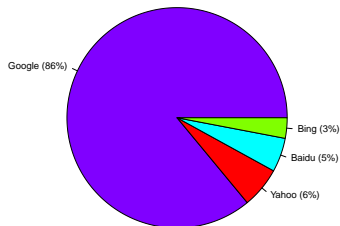


El éxito de Google: Aplicar el principio KISS⁴

Evolución de Yahoo vs. Google:



Cuota de mercado de los motores de búsqueda:



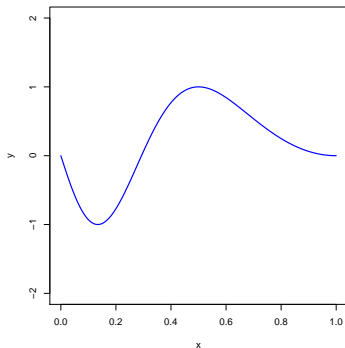
⁴En estadística este se conoce como **Principio de Parsimonia**.

El problema del modelado

Considere la función

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

cuyo gráfico es dado por:

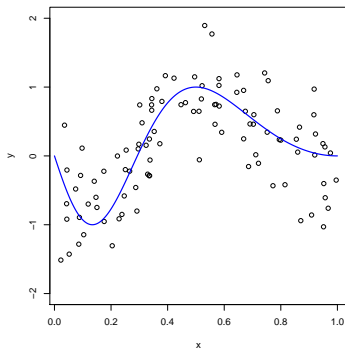


El problema del modelado

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 $\sigma = 1/2$,

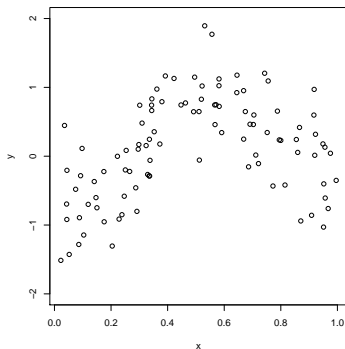


El problema del modelado

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:

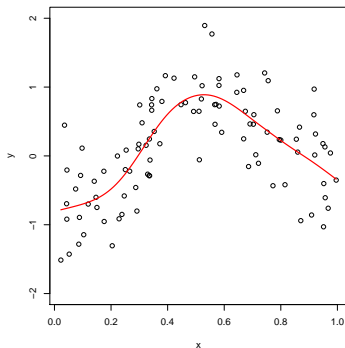


El problema del modelado

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

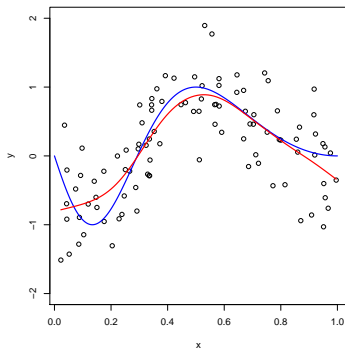


El problema del modelado

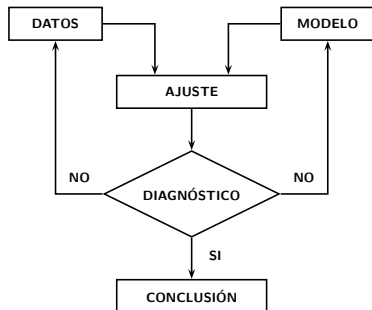
En Estadística se estudia teóricamente, la “bondad del modelo” comparando

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

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.

Análisis de **Sensibilidad**.

Comuniqué sus resultados!

