

Introducción

Estadística Computacional

Juan Zamora Osorio
juan.zamora@pucv.cl

Instituto de Estadística
Pontificia Universidad Católica de Valparaíso

28 de agosto de 2023



PONTIFICIA
UNIVERSIDAD
CATÓLICA DE
VALPARAÍSO

Probabilidades y estadística

Filosofía / Epistemología

- ▶ ¿Lo que sucede en el mundo depende del azar?
- ▶ ¿Qué es el azar?

Ejemplos

- ▶ Un juego de azar.
- ▶ Máquinas de azar.
- ▶ Cuánto demora en responder una consulta web.
- ▶ El clima de mañana.
- ▶ Cuántas unidades de un producto se venden a través de un sitio web.
- ▶ Si un estudiante llega a la clase o no.
- ▶ Foto astronómica de estrellas lejanas.
- ▶ Partículas observadas en acelerador de partículas.

Probabilidades y estadística

Filosofía / Epistemología

- ▶ ¿Lo que sucede en el mundo depende del azar?
- ▶ ¿Qué es el azar?

Pregunta

- ▶ ¿Cuál es la diferencia entre *azar*, *aleatorio* y *estocástico*?

Probabilidades y estadística

Pregunta

- ▶ ¿Cuál es la diferencia entre *azar*, *aleatorio* y *estocástico*?

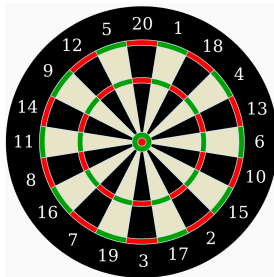
Etimología

- ▶ *Azar* (castellano): De *azzahr*, juego de dados, y éste del árabe *zahr*, usado como *dado* y que significa literalmente *flores*.
- ▶ *Aleatorio* (latín): De *alea*, suerte, usado como sinónimo de juego de azar.
- ▶ *Estocástico* (griego): De *stokhastikos*, apuntar a un blanco, conjeturar, a su vez de *stokhos*, el objetivo apuntado.

Probabilidades y estadística

Pregunta

- ¿Cuál es la diferencia entre *azar*, *aleatorio* y *estocástico*?



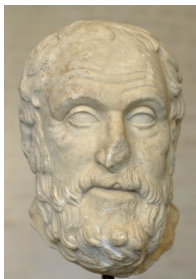
Entre dados y dardos

- La diferencia no es solo una letra *r*, sino una concepción de mundo diferente.

Probabilidades

Origen: Carneades, siglo II a.C.

- ▶ Escéptico: No es posible conocer algo absolutamente.
- ▶ Propone que toda decisión (jurídica / política) tiene incertidumbre.
- ▶ Para funcionar de manera práctica, asignar un valor de verdad a las afirmaciones, asociado a seguridad que sujeto tiene de afirmación.



Probabilidades

Teoría de Probabilidades

- ▶ Se interpreta como la modelación matemática del azar, ella debe dar cuenta del conocimiento acumulado por sobre los fenómenos del azar y sus leyes. La noción de probabilidad está subordinada a una determinada aproximación a los fenómenos del azar.

Actual: Andréi Kolmogórov (1930)

- ▶ Teoría basada en el cálculo matemático de medida, que explica al menos las teorías de las leyes del azar en diferentes contextos.



Probabilidades – Leyes del azar

Ley de los grandes números

- ▶ Modelo matemático para comportamiento del promedio, permite construir visión frecuentista de la probabilidad.

Ley del comportamiento de las fluctuaciones

- ▶ Estudio de las fluctuaciones de las pequeñas variaciones y sus correspondientes modelos matemáticos referidos a diferentes versiones del teorema del límite central.

Ley de la complejidad

- ▶ La complejidad de todo sistema dinámico aumenta en el curso de su evolución (Boltzmann).

¿Qué es la estadística?

Varias definiciones

- ▶ Cómo opinar y tomar decisiones bajo la presencia de incertidumbre.
 - ▶ Ciencia y arte de tomar decisiones basadas en evidencia cuantitativa.
 - ▶ Colección de métodos que nos ayudan a describir, resumir, interpretar y analizar datos.
-
- ▶ Sin incertidumbre, ¿hay necesidad de métodos estadísticos?

¿Qué es la estadística?

Historia

- ▶ Proviene de la ciencia política de recolectar datos para describir poblaciones, negocios, etc. para administrar un estado.
- ▶ Siglo XIX se amplía para diseñar y analizar experimentos en agricultura.
- ▶ Siglo XX se amplía a industrias.
- ▶ Siglo XXI: ciencia de datos.

¿Dónde se utiliza?

- ▶ Casi en todas las áreas del conocimiento que recolectan e interpretan datos.
- ▶ Transformar datos en información.

Estadística y computación

Explosión

- ▶ Volumen
- ▶ Bit
- ▶ costo
- ▶ Le

FIGURE 10. AVERAGE \$/GB FOR HDDS

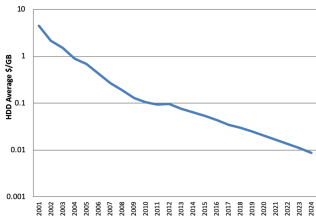
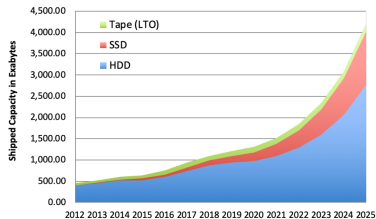


FIGURE 11. CAPACITY SHIPMENTS FOR LTO TAPE, SSDS AND HDDS



Estadística y computación

Computación

- ▶ Procesar dichos datos y extraer valor es fundamental.
- ▶ Volumen requiere algoritmos y competencias en computación.
- ▶ Algoritmos probabilísticos.

Gran diferencia

- ▶ Aprender a vivir con datos que tienen errores.
- ▶ *El conocimiento estadístico es fundamental.*

Estadística

Matemáticas

- ▶ En matemáticas muchos problemas tienen una respuesta única y en que estamos todos de acuerdo.
- ▶ ¡Se requieren supuestos que pueden llegar a conclusiones diferentes!

Otras disciplinas – “Arte”

- ▶ Parte del trabajo es identificar el problema para poder usar las herramientas matemáticas.
- ▶ Se requiere conocer el área de estudio.
- ▶ Experiencia y creatividad.

Comunicación

- ▶ Cómo presentar resultados a clientes.
- ▶ Cómo comunicar incertidumbre.
- ▶ Cómo comunicar riesgo.

Ciencia de datos (*data science*)

Intersección multidisciplinaria

- ▶ Computación y aprendizaje automático (*Machine Learning*).
- ▶ Estadística.
- ▶ Dominio de aplicación.

Análisis estadístico

Partes

- ▶ Recolección de datos.
- ▶ Administración de los datos.
- ▶ Aplicación de procedimientos estadísticos.
- ▶ Interpretación de los resultados.

Este curso

- ▶ Nos concentraremos en las dos últimas.
- ▶ *¡Todas son importantes!*

Análisis estadístico

Tipos

- ▶ Descriptivo.
- ▶ Predictivo.
- ▶ Causalidad.

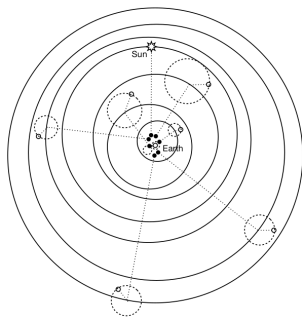


Análisis estadístico

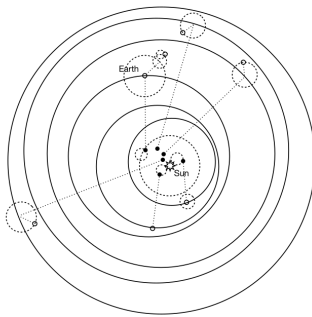
Tipos

- ▶ Descriptivo.
- ▶ Predictivo.
- ▶ Causalidad.

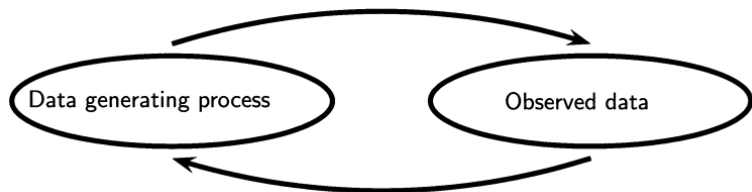
Ptolemaic Model



Copernican Model



Probabilidades e inferencia estadística



Probabilidades

- ¿Dado un proceso que genera datos, cuáles son las propiedades que observaremos?

Inferencia estadística

- ¿Dadas las observaciones, qué podemos decir sobre el proceso que genera los datos?

Enlaces útiles – Conjuntos de datos

Internacionales

- ▶ Kaggle: <https://www.kaggle.com/datasets>.
- ▶ Google: <https://datasetsearch.research.google.com/>.
- ▶ Google: <https://www.google.com/publicdata>.
- ▶ UCI: <https://archive.ics.uci.edu/ml/datasets.php>.

Enlaces útiles – Conjuntos de datos

Nacionales

- ▶ Instituto Nacional de Estadísticas: <https://ine.cl/>.
- ▶ Dirección Meteorológica de Chile:
<https://climatologia.meteochile.gob.cl/>.
- ▶ Banco Central: <https://www.bcentral.cl/inicio>.
- ▶ Centro Sismológico Nacional:
<https://www.sismologia.cl/index.html>.
- ▶ Infraestructura de Datos Geoespaciales de Chile:
<https://www.ide.cl/>.
- ▶ Estadísticas Servicio de Impuestos Internos:
<https://www.sii.cl/destacados/ogp/index.html>.
- ▶ Biblioteca del Congreso Nacional:
<https://www.bcn.cl/leychile/>.
- ▶ Coordinador Eléctrico Nacional: <https://www.coordinador.cl/>.
- ▶ Energía Maps: <https://energiamaps.cne.cl/>.

- ▶ **Information theory, inference and learning algorithms.** Cambridge university press, 2003. David J. C. MacKay.
- ▶ **Probability & Statistics for Engineers & Scientists.** Pearson Education, 2012. Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers, Keying Ye.
- ▶ **Probability & Statistics for Engineering and the Sciences.** Cengage Learning, 2015. Jay L. Devore.
- ▶ **Introduction to Statistics and Data Analysis.** Springer, 2016. Christian Heumann, Michael Schomaker Shalabh.
- ▶ **Probability and Statistics for Computer Science.** UK: Springer International Publishing, 2018. David Forsyth.

Enlaces útiles

- ▶ **Statistical rethinking: A Bayesian course with examples in R and Stan.** Chapman and Hall/CRC, 2018. Richard McElreath.
- ▶ **Applied Multivariate Statistical Analysis.** Springer Nature, 2019. Wolfgang Karl Härdle, Léopold Simar.

Otros cursos

- ▶ **Probabilidad y Estadística USM.** Ronny Vallejos.
https://www.youtube.com/playlist?list=PLRdsr8w_wLNzYYSYP6bvf1p30mo27X9q-.
- ▶ **Pensamiento estadístico.** Felipe Bravo-Marquez.
<https://github.com/dccuchile/CC6104>.