

## Estadística para Simulación (B-2017)

Profesores: Magdiel Ablan B. (mablan@ula.ve, mablan@gmail.com)  
Paolo Ramoni Perazzi (rpaolo@ula.ve, rpaolo1967@gmail.com)  
Centro de Simulación y Modelos (CESIMO)  
Tel. 2403002

### Introducción

Una fase muy importante de la simulación consiste en realizar experimentos con un modelo de un sistema que posee una o más variables estocásticas o aleatorias a lo largo del tiempo. Este curso pretende dar respuesta a varias cuestiones relacionadas con esta tarea como: ¿Cómo modelizar estas variables? ¿Cuál es el alcance de la información que se deriva de estos modelos aleatorios? ¿Cómo se puede usar el computador para analizar los datos de entrada o salida de una simulación?

### Objetivos

El curso hace énfasis en los métodos estadísticos computacionales y en el uso del software estadístico apropiado para manejar grandes cantidades de información. El enfoque del curso es pragmático: los conceptos son presentados pero no se hace énfasis en el tratamiento matemático riguroso de los mismos sino en el procesamiento e interpretación de los datos.

Específicamente los objetivos de este curso son:

1. Aprender algunas de las herramientas de estadística computacional más usadas en la simulación: Métodos de Monte-Carlo, Generación de V.A, Análisis de Datos.
2. Repasar algunos de los conceptos fundamentales de la estadística
3. Aprender R, una de las herramientas computacionales comúnmente usadas para cálculo estadístico y análisis de datos

### Contenido

1. **Unidad I: Introducción al curso y al lenguaje R.** El objetivo de esta unidad es introducir al estudiante en el uso del lenguaje R. Los tópicos cubiertos incluyen:
  - El ambiente de trabajo en R
  - Estructuras de datos en R: Vectores, arreglos, marco de datos, listas.
  - Funciones en R.
  - Lectura de datos
  - Gráficos.

- Como instalar nuevos paquetes
  - Programación básica en R
2. **Unidad II: Repaso de Estadística.** El objetivo de esta unidad es revisar rápidamente algunos de los principales conceptos de estadística que se cubren en los cursos de estadística de primer nivel e ilustrar el uso del R como herramienta estadística. Los tópicos cubiertos incluyen:
- Variables aleatorias. Función de densidad y de probabilidad acumulada. Propiedades y momentos
  - Variables aleatorias continuas y discretas: las más comunes
  - Inferencia estadística. Distribuciones muestrales, teoremas fundamentales
  - Intervalos de confianza y pruebas de hipótesis
  - Estadísticos y sus métodos de estimación
3. **Unidad III: Visualización de datos.** En esta unidad se revisan varios métodos de visualización de datos que pueden resultar útiles en el análisis de datos de entrada o salida de una simulación y como paso previo a otros análisis. Entre los tópicos cubiertos se encuentran:
- Librerías o paquetes gráficos en R.
  - Visualización de varios paneles
  - Gráficos de superficies y dispersión de puntos en 3D.
  - Gráficos de Contorno
  - Otras representaciones 2D de los datos
  - Enfoques alternativos de visualización.
4. **Unidad IV: Métodos para la generación de variables aleatorias.** En esta unidad se revisan varios de los enfoques generales para simular variables aleatorias. Específicamente los tópicos cubiertos son:
- El método de la transformada inversa
  - El método de aceptación y rechazo
  - Métodos de transformación
  - Simulación de sumas y mezclas de variables aleatorias
  - Simulación de procesos estocásticos
5. **Unidad V: Métodos de Monte Carlo.** En esta unidad se aprende las bases del método de Monte Carlo y varias de sus aplicaciones en la estadística. Específicamente se revisarán los temas siguientes:
- Integración de Monte Carlo
  - Reducción de la varianza y variables antitéticas
  - Métodos de y Monte Carlo en inferencia: estimación y prueba de hipótesis
  - Pruebas de Permutación
6. **Unidad VI: Bootstrap y Jackknife.** En esta unidad se introducen los métodos de remuestreo (bootstrap) y jackknife y varias de sus aplicaciones. Los temas considerados son:

- Bootstrap
- Jackknife
- Intervalos de confianza usando bootstrap
- Pruebas de permutación

7. **Unidad VII: Otros tópicos:** Una pequeña introducción a otros tópicos como optimización, simulación de cadenas de Markov Monte Carlo, estimación de densidad u otro.

Los libro texto fundamentales son : Rizzo (2008) y Owen y otros (2009)

## Metodología

*"Nunca enseñe a mis estudiantes. Sólo intento proporcionar las condiciones en las que ellos puedan aprender"*

Albert Einstein

*"Lo que escucho olvido, lo que veo recuerdo, lo que hago comprendo"*

Confucio

Estas dos frases resumen la metodología de enseñanza en este curso. Si bien tendremos clases magistrales y el material de las clases estará disponible desde el principio, donde realmente se aprende no es clase, si no cuando se realizan los ejercicios y tareas.

## Evaluación

- Tareas o Mini-proyectos (60%): A ser entregadas en la fecha correspondiente. Evaluadas sobre 15 para retardos desde un día hasta una semana y sobre 10 para retardos de hasta dos semanas.
- Pendiente a ponernos de acuerdo (20%).
- Examen (20%): Hacia el final del curso. Similar a la tareas pero corto para que pueda ser completado en el espacio de la clase

## Recursos

Estamos en el medio de una explosión de conocimientos y recursos relacionados con el análisis de datos, la estadística y el lenguaje R. Hoy en día existe demasiada información. Sólo con referencia a R, editoriales importantes como Springer tiene una serie de libros dedicada a diversos aspectos de R, *Use-R* con 47 títulos la última vez que miré. Chapman and Hall tiene otra

colección similar, *The R series* con más de veinte títulos. Así, la siguiente lista no es para exhaustiva, como un somero intento de búsqueda en Google lo revelaría.

## Software

- El software estadístico R. El software que utilizaremos durante el curso. <http://cran.r-project.org>. El punto de partida a muchos recursos interesantes.
- R studio. Una interfaz para R, en mi opinión, la mejor: <http://www.rstudio.com/>

## Cursos en línea

Coursera (<https://www.coursera.org/>), EDX (<https://www.edx.org/>) y Udacity (<https://www.udacity.com/>) todos tienen cursos sobre estadística y análisis de datos que son afines o complementarios a este curso. Dictados por excelentes profesores de prestigiosas universidades y gratis. Por mencionar sólo uno especialmente relevante:

Computing for Data Analysis (<https://class.coursera.org/compdata-003/class/index.>) Roger Peng. de Johns Hopkin Bloomberg School of Public Health.

## Libros

Albert Jim y María Rizzo. 2012. *R by example*. Springer

Jones Owen, Robert Maillardet, and Andrew Robinson. 2014. *Introduction to Scientific Programming and Simulation Using R*. Chapman and Hall/CRC.

Law A. M. and W. D. Kelton. 1982. *Simulation Modelling and Analysis*. McGraw-Hill Book Co.,

Nolan Deborah and Lang Temple, Duncan. 2015. *Data Science in R : Case Studies Approach to Computational Reasoning and Problem Solving..* CRC Press.

R Development Core Team. 2008. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0.

Rizzo, María. 2008. *Statistical Computing with R*. Chapman & Hall/CRC,