

Análisis Masivo de Datos I

MSc. Alvaro Limber Chirino

Unidades de aprendizaje

1. Introducción al Big Data
2. Fuentes de datos
3. Gramática de datos
4. Modelado con software con MapReduce
5. Muestreo y filtrado

Evaluación

Formal

- 5 Laboratorios, 10pts/cu: 50%
- Proyecto Dashboard: 20%
- 2 exámenes teóricos, 15pts/cu: 30

Extra

- Poster estadístico: 10%

Referencias

- Erl, Thomas, Wajid Khattak, and Paul Buhler. 2016. *Big Data Fundamentals Concepts, Drivers & Techniques*. Iacus, Stefano M. 2015. *Automated Data Collection with r - a Practical Guide to Web Scraping and Text Mining*. *Journal of Statistical Software*. Vol. 68. <https://doi.org/10.18637/jss.v068.b03>.
- . 2017. *Big Data and Social Science - a Practical Guide to Methods and Tools*. *Journal of Statistical Software*. Vol. 78. <https://doi.org/10.18637/jss.v078.b02>.
- Luraschi, J., K. Kuo, E. Ruiz, and an O'Reilly Media Company Safari. 2019. *Mastering Spark with r*. O'Reilly Media, Incorporated. <https://therinspark.com/>.
- Wickham, H., and G. Grolemund. 2017. *R for Data Science*. O'Reilly Media. <https://r4ds.had.co.nz/>.

Diagnóstico

- Importar la base de datos de la encuesta de demografía y salud
- Explorar las variables de altura, peso para las mujeres de 15 años o más
- Construir el Índice de Masa Corporal (IMC)
- Estimar e interpretar el modelo lineal:

$$IMC_i = \beta_0 + \beta_1 Peso_i + \beta_2 Altura_i + \epsilon_i$$