

# Taller de Softwares: Python, R y Latex

Grupo Lambda 2020

## I. PROFESOR

### **Mg. Heber Julio Baldeón Paucar**

Coordinador de Asociaciones Público Privadas, Subsidios Cruzados y Métodos Cuantitativos en la Gerencia de Regulación Tarifaria de la SUNASS. Economista de la UNMSM, Máster en Finanzas Cuantitativas en la Universidad de Alcalá de Henares, Madrid – España y certificación internacional: Certified in Risk Management (CRM) del IIPER Institute y Chartered Risk Analyst (CRA) del American Academy of Financial Management. Además, ha cursado el 56 Curso de Extensión Universitaria de Economía Avanzada del BCRP y especializaciones en el FMI, Banco Mundial, ESAN, UP, PUCP y UNMSM. Especialista en finanzas, econometría aplicada, machine learning y gestión de riesgos.

## II. OBJETIVOS

- El objetivo del curso busca que el participante utilice el lenguaje de programación Python y R con el fin de manipular datos, generar análisis estadísticos y representación gráfica. Así como, preparar documentos y presentación en el editor Latex.
- El curso está dirigido a estudiantes y egresados de Economía, Finanzas, Ingeniería, Estadística y carreras afines, interesados en conocer y aprender el lenguaje de programación de Python y R con múltiples aplicaciones en Estadística.

## III. METODOLOGÍA

El curso consta de 28 horas lectivas. Las sesiones consisten en exposiciones por parte del docente y la participación activa de los alumnos mediante el desarrollo de aplicación en cada clase.

## IV. REQUISITOS

El participante debe tener conocimientos básicos de estadística.

## V. CONTENIDO

### **Sesión I y II: Editor Latex para ciencias sociales con TexStudio (8 horas)**

Configuración del TexStudio y uso de paquetes.

Estructura y edición de documento de textos y científicos en Latex

Compilación de documentos.

Matemáticas, figuras y tablas en Latex, y bibliografía automática con BibTex.

Presentaciones con Beamer en Latex

Interacción con otros softwares (Stata, R, otros).

### **Sesión III: Introducción a la programación en Python y Spyder (4 horas)**

Instalación de Python y Anaconda.

Importación de librerías Pandas y NumPy (Numerical Python).

Creación y carga de bases de datos en formato csv, xls, xlsx, tex, HTML.

Categorizando información cualitativa.

Merge, reshape, clean y transformación de variables.

Missing data, duplicados, reescalando, normalización (estandarización).

Detección de outliers.

### **Sesión IV: Visualización de datos con Python y Spyder (4 horas)**

Visualización de datos con el paquete ggplot2.

Gráficas de dispersión, líneas, barras, distribuciones, mapas de color.

Gráficas en 3D y animaciones, textos e histogramas en 3D.

### **Sesión V: Introducción a la programación y manejo de Bases de Datos con RStudio (4 horas)**

R y el R Studio, paseo por el interfaz del R Studio.

Creación y carga de bases de datos en formato csv, xls, xlsx, tex, HTML.

Limpieza de datos: datos NA, duplicados, reescalando, normalización (estandarización)

Categorizando información numérica

### **Sesión VI: Análisis exploratorio y estadístico de datos con RStudio (4 horas)**

Combinación y separación de datos

Detección de outliers

Creación de funciones en R

Análisis de causalidad

### **Sesión VII: Visualización de datos con RStudio (4 horas)**

El paquete ggplot2

Gráficas de dispersión, líneas, barras, distribuciones, treemaps, mapas de color

Gráficas 3D y animaciones, textos e histogramas en 3D.

Medidas estadísticas y análisis estadística