

Curso online:
Análisis de datos y Econometría Aplicada en R
Básico – Intermedio
Grupo Lambda
2020

1. Profesor

Mg. Heber Julio Baldeón Paucar

Coordinador de Asociaciones Público Privadas (APP), subsidios cruzados y métodos cuantitativos en la Gerencia de Regulación Tarifaria de la SUNASS. Economista por la UNMSM y Máster en Finanzas Cuantitativas en la Universidad de Alcalá, Madrid - España. Además, aprobó el 56 Curso de Extensión Universitaria de Economía Avanzada del BCRP y cuenta con especializaciones en el FMI, Banco Mundial, ESAN, UP, PUCP y UNMSM. Su ámbito de interés es la econometría aplicada, *machine learning*, finanzas cuantitativas y gestión de riesgos.

2. Descripción del curso

R es la lengua franca de la estadística. Es un lenguaje de programación con gran cantidad de paquetes estadísticos y desarrollo de nuevas técnicas estadísticas, lo que te permite realizar cualquier análisis y visualización de datos de manera sencilla y rápida, con pocas líneas de código.

R es un lenguaje Orientado a Objetos, un lenguaje interpretado (como Java) y no compilado (como Fortran o Pascal), y multiplataforma. Así, R es un ambiente de programación estable, confiable y muy flexible que pueden extenderse fácilmente mediante paquetes o librerías o definiendo nuestras propias funciones. Además, es gratuito y de código abierto, con una amplia comunidad colaborativa.

Por eso, este curso de guiará desde las estructuras nativas de bases de datos, su tratamiento y visualización, hasta la implementación de modelos econométricos lineales, no lineales y de series de tiempo univariados y multivariados.

3. Objetivos del curso

- Proporcionar al estudiante los fundamentos básicos intermedios de programación y manejo de bases de datos en el software libre R y R Studio.
- Proporcionar al estudiante las herramientas cuantitativas básicas para estimación de modelos lineales para datos de corte transversal y series de tiempo.
- Proporcionar al estudiante la utilidad del software R en aplicaciones de Econometría.
- Uso intensivo de bases de datos como el ENAHO y BCRP.

4. Dirigido a:

- Estudiantes de pre-grado o egresados de economía, estadística e ingeniería que busquen iniciar su proceso de aprendizaje del lenguaje R, lengua franca de la estadística.

5. Requisitos

Conocimiento básico de estadística y álgebra lineal.

6. Metodología

- El curso consiste en 16 horas lectivas. Las sesiones consistirán en exposiciones del Docente y resolución de casos aplicados.

- Se otorgará un certificado a nombre de la institución por la participación en el curso. Además, de un trabajo integrador opcional para la emisión de un certificado de aprobación a nombre de la institución.

7. Contenido

7.1 Estructura de datos e introducción a la programación (2 horas)

- R y el R Studio, paseo por el interfaz del R Studio.
- Estructuras nativas de datos: Arrays, matrix, factor, list, data.frame.
- Aritmética en RStudio
- Algebra lineal en RStudio: matrices, inversa y determinante, descomposición de matrices, otros.
- Calculo en RStudio: derivadas e integrales.
- Instalación e importación de librerías
- Creación y carga de bases de datos en formato csv, xlsx, tex, HTML, json.
- Creación de estructuras de control: condicional, loops y funciones.

7.2 Tratamiento y visualización de datos (4 horas)

- Merge, append, reshape de datos.
- Transformación de variables y categorizando información cualitativa.
- Tablas de resumen y filtros.
- Reescalando, normalización (estandarización).
- Missing data, duplicados, detección de outliers.
- Visualización de datos con el paquete ggplot2.
- Gráficas de dispersión, líneas, barras, distribuciones, mapas de color.
- Gráficas 3D y animaciones, textos e histogramas en 3D.

7.3. Análisis de Regresión (6 horas)

- Estimadores por Mínimos Cuadrados Ordinarios y Máxima Verosimilitud.
- Regresión lineal simple, regresión lineal múltiple y regresión polinomial. Supuestos y violación de supuestos.
- Regresión logística. Modelos Probit, Poisson, Tobit y multinomiales.
- Evaluación del performance del modelo de regresión y matriz de confusión.

7.4. Análisis de Series de Tiempo (4 horas)

- Secuencia y Operaciones con fechas.
- Descomposición y filtros para series de tiempo.
- Suavizamiento y predicción con el método de Holt-Winters.
- Modelos ARMA, ARIMA y SARIMA.
- Series de tiempo multivariadas: VAR, VEC, SVAR y SVEC
- Proyección de series de tiempo por Redes Neuronales.

Lima, junio de 2020