

Introducción al programa estadístico R
CURSO DE POSGRADO
PEDECIBA

Coordinadoras y docentes: Carolina Crisci y Angel Segura, CURE

Docente participante: Juan Manuel Piñeiro

Institución proponente: CURE

Público objetivo: Estudiantes de posgrado. En casos justificados se aceptarán estudiantes avanzadas de licenciatura.

Carga horaria:

El curso se dictará en modalidad híbrida con instancias presenciales y encuentros virtuales sincrónicos. Se espera que el estudiante le dedique al curso al menos 4 horas semanales, aparte de los encuentros. Se disponibilizará el material en la plataforma EVA estructurado por clase. En total, se prevé una carga horaria de 30 horas de encuentros distribuidas a lo largo del curso de la siguiente manera: dos encuentros por semana de 2 horas y media. Uno de estos encuentros será virtual (martes por zoom) y el otro presencial (jueves, CURE, Sede Rocha, salón 2). Se requiere la participación **obligatoria en al menos un encuentro presencial por módulo**, aunque se recomienda la participación en todos los encuentros presenciales.

Total: 30 horas

Fecha de inicio: 20 de agosto

Fecha de finalización: 26 de setiembre

Días del curso: martes y jueves

Horario: 9:30 a 12:00

Cupo máximo: 25

Modalidad del curso

Modalidad híbrida. Los martes los encuentros serán virtuales por Zoom, los jueves habrá una instancia presencial en el CURE, sede Rocha. Se utilizará la plataforma EVA.

Materiales: prácticos en R (trabajaremos con RStudio) que serán proporcionados en formato

html, dicho formato se genera a partir de un tipo de archivo de RStudio denominado R Markdown (.Rmd). Este formato es didáctico y muy amigable para las y los estudiantes ya que muestra el código que debe introducirse en el programa y debajo de las distintas secciones de código se muestra el resultado del mismo (estadísticos, tablas, figuras).

Bases de datos

Videos sobre contenidos específicos.

Bibliografía en español e inglés (pdf, páginas web).

Objetivos del curso

Introducir al/la participante en el uso del programa R, brindándole conocimientos para levantar y exportar datos, manipularlos, calcular estadísticos, generar gráficos, realizar test estadísticos de uso frecuente y programar a nivel básico. Brindar herramientas, conceptos y características teóricas/análisis de distribuciones en R (generación de números pseudoaleatorios con distribuciones determinadas, cálculo de la función de densidad y de probabilidad de una variable, cálculo de la función cuantil, pruebas de bondad de ajuste). Selección de modelos lineales por verosimilitud (lm y glm). Determinación y comparación de la capacidad de predictiva de modelos (validación cruzada, error por muestra test, RMSE). Presentación de resultados de modelos estadísticos en gráficos y tablas. Como presentar los modelos en un informe escrito. No se profundizará en el trasfondo teórico de los métodos a aplicar. Vale aclarar que el curso no es un curso de estadística por lo que todo lo referido al análisis estadístico de datos será dictado sin profundizar en el trasfondo teórico.

Contenidos

Módulo 1 (2 semanas): Instalación del R, primera aproximación al programa, reconocimiento de sus componentes y funcionalidades. Potencialidades. ¿Qué es un “objeto”? ¿Qué tipos de objetos hay?, Operaciones sencillas, Generación de datos (mediante las funciones `c()`, `rep()`, `seq()`, `rnorm()`, etc.).

Estadísticos: media, varianza, desvío estándar, error estándar, cuantiles, mínimo, máximo, rango. Tablas simples y de doble entrada. Levantar datos desde archivos “.csv”. Exploración básica de datos mediante la función `summary()`,

Manipulación de datos (quitar/agregar filas y columnas, cambiar nombres de variables, ordenar datos, recodificar variables, transformar variables continuas en factores, etc.), datos faltantes en R, ¿Cómo instalar y cargar paquetes?, ¿Cómo exportar datos?, ¿Cómo obtener ayuda? Cambios en el aspecto de R. Otras funciones básicas: `dim()`, `nrow()`, `ncol()`, `which()`, `sample()`, `colSums()`, `rowSums()`, etc.

Gráficos con `ggplot`, funciones (bucles *for*, funciones de la familia *apply*, funciones definidas por el usuario). Otros análisis y gráficos: Correlación entre dos variables (correlación de Pearson y

de Spearman), matriz de correlaciones (varias variables), gráficos para visualizar la matriz de correlación. Básico de modelos lineales, comentarios sobre las potencialidades del RStudio para la publicación y divulgación de materiales didácticos.

Módulo 2 (2 semanas): Distribuciones de probabilidad. Breve repaso de conceptos y características teóricas/análisis de distribuciones en R (generación de números pseudoaleatorios con distribuciones específicas, estimación de la función de densidad y de probabilidad de una variable, cálculo de la función cuantil, pruebas de bondad de ajuste).

Distribuciones continuas que se verán: Uniforme, Normal, t de Student, Chi-cuadrado de Pearson, Gamma, Exponencial, F de Fisher-Snedecor, Weibull. Distribuciones discretas: Binomial, Binomial Negativa, Poisson, Hipergeométrica.

Módulo 3 (2 semanas): Selección de modelos (Verosimilitud y AIC). Selección de modelos lineales por verosimilitud (lm y glm). Determinación y comparación de la capacidad de predictiva de modelos (validación cruzada, error por muestra test, RMSE). Presentación de resultados de modelos estadísticos en gráficos y tablas. Como presentar los modelos en un informe escrito. No se profundizará en el trasfondo teórico de los métodos a aplicar.

Nota: Previo al comienzo del curso se disponibilizará en EVA un tutorial breve para instalación del programa R y RStudio e información sobre los componentes básicos del programa. Es importante que las y los estudiantes tengan R y RStudio instalado al momento de comenzar el curso.

Evaluación

Examen escrito individual que consistirá en una serie de ejercicios en R de aplicación de lo visto en el curso. Se realizará en EVA.

Bibliografía de referencia

Karline Soetaert, Peter M.J. Herman 2009. A Practical Guide to Ecological Modelling. Using R as a Simulation Platform. ISBN: 978-1-4020-8624-3
Benjamin M. Bolker 2008. Ecological Models and Data in R. <https://doi.org/10.1515/9781400840908>.
Canavos, J. 1988. Probabilidad y Estadística, Aplicaciones y Métodos. Naucalpan de Juárez, McGraw-Hill
Crawley, Michael J. The R Book. Chichester, England: Wiley, 2007. Print.
R para Ciencia de Datos. Garrett Grolemond & Hadley Wickha. Version en línea en español:
<https://es.r4ds.hadley.nz/index.html>
Otros links útiles:
<https://bookdown.org/jboscomendoza/r-principiantes4/>
<https://www.r-project.org/>
<https://rstudio.com/about/>