

## ACTIVIDAD DE SERVICIO TÉCNICO

### “INTRODUCCIÓN A R”

**Incluida en el proyecto de extensión docente ED-3480**

---

**Unidad académica:** Centro de Investigación en Matemática Pura y Aplicada, CIMPA.

**Vigencia de la actividad:** 6 semanas, fechas por definir.

Con esta actividad, se quiere ofrecer un curso de 4 semanas sobre los aspectos básicos del lenguaje estadístico R y sus paquetes orientados al manejo de datos y análisis.

**Objetivo general:** el curso busca desarrollar habilidades en los participantes para el análisis de datos usando el software libre R, de manera que se tenga un acercamiento a las técnicas estadísticas básicas.

**Alcance:** En respuesta a las necesidades que muestran las empresas e instituciones en aumentar su conocimiento en el uso de herramientas numéricas eficaces para el manejo de datos y su análisis, este curso presenta una introducción práctica de elementos de estadística descriptiva aplicada en R para usos en el sector empresarial e institucional.

**Antecedentes:** Como consecuencia de la creciente tecnificación de nuestra sociedad, desde hace varios años se ha observado un aumento en el flujo de solicitudes de asesorías, consultorías y capacitaciones por parte del sector externo de la Universidad de Costa Rica, en temas muy diversos relacionados con el área de las matemáticas aplicadas. Por ejemplo, nuestro centro frecuentemente recibe solicitudes de estudios actuariales de planes de pensiones y jubilaciones, cursos de capacitación en gestión de riesgo financiero, asesorías en análisis de datos y modelación matemática, análisis de estructuras salariales, peritazgo y arbitraje matemático en juicios de toda naturaleza. Esto ha sido acompañado por un aumento en el interés del sector productivo en contratar personal con una sólida formación matemática y en capacitar en matemática al personal que ya tienen contratado para así mejorar su desempeño laboral.

**Justificación:** Debido al interés mostrado por algunas instituciones y empresas, se realizará una capacitación sobre "Uso del paquete en R", el cual cubrirá aspectos prácticos en el uso del paquete R para generar insumos que sirvan de apoyo en la toma de decisiones más acertadas en su quehacer.

**Descripción:** El curso está organizado en sesiones semanales durante 4 semanas. Los docentes darán los materiales de estudio, impartirán las clases y mantendrán contacto en la semana para evacuar consultas de los participantes. Los participantes se comprometen a participar en las clases con su asistencia y su participación activa, la cual supone el estudio de las metodologías, la realización de los ejercicios y la formulación de preguntas atinentes al contenido del curso.

**Requisitos:** no se requiere ningún conocimiento previo de R.

**Perfil de participantes:** trabajar en el sector productivo nacional (público o privado) y tener familiaridad con datos numéricos repetitivos; tener una formación académica en que se haya formado en el uso de tablas de datos, manejo de datos estadísticos, alguna familiaridad con software para el manejo numérico de datos, entre ellos hojas electrónicas, y una buena disposición para aprender y participar del curso. Diversas carreras forman en estas habilidades, entre ellas: administración de negocios, administración pública, banca y finanzas, estadística, economía, ciencias actuariales, dirección de empresas, diversas ingenierías, ciencias agroalimentarias, algunas carreras en el sector salud, etc.

**Metodología:** los contenidos serán motivados a partir de datos y la necesidad de hacer un tratamiento a los mismos. Luego se verán distintas herramientas en R para el manejo y visualización de datos. Se requiere que los participantes hagan constantemente trabajo práctico sobre tablas de datos y reflexionen sobre el uso apropiado de las técnicas.

**Modalidad:** dadas las restricciones actuales derivadas de la pandemia, se trabajará de manera remota usando la plataforma UCR Global, el sistema de reuniones virtuales *Zoom* y otras herramientas para el trabajo remoto.

**Horario de clase:** día de la semana a definir, de 8:30 a.m. a 12:00 m.d. Entre semana, cada participante debe dedicar unas 17 horas para la práctica semanal.

**Certificación:** Al final del curso se entregará un certificado de (participación/aprovechamiento) por 80 horas.

### **Contenidos:**

- Clase 1. Introducción
  - Contexto y preliminares: ¿Por qué es importante R? Uso de RStudio. Instalación de paquetes

- Nociones de R: Tipos de objetos. Ciclos y condicionales. Funciones y argumentos. Manejo de vectores. Manejo de NA's (missing values).
- Manejo de datos: Estructura del *data.frame*. Lectura de datos (*base, readr, fread*). Manejo de los elementos de base datos. Manejo de fechas y caracteres.
- Cuestionario: tipo de datos y gráficos de interés para los participantes, para ver en la clase 4.
- Clase 2. Manipulación de datos
  - Limpieza de datos y transformaciones (*dplyr*): Selección de columnas o filas. Encadenamiento de comandos. Eliminación de columnas. Combinación de múltiples *data.frames*; paquete *janitor*
  - Manejo de factores (*forcats*), cadenas de caracteres (*stringr*), formatos de fechas (*lubridate*)
  - Definición de datos ordenados: estructura de tablas en formato largo y ancho (*tidyr*).
  - Operaciones agrupadas (*dplyr*).
- Clase 3. Visualización básica
  - Visualización (*ggplot2*): Sintaxis y estructura de un *ggplot*.
  - Creación de histogramas, gráficos de caja. Gráficos bajo datos categóricos.
  - Gráficos relacionales: gráficos lineales, diagramas de dispersión, matrices de gráficos.
  - Cambio de elementos estéticos (color, formas, tamaños).
  - Creación de múltiples figuras al mismo tiempo.
- Clase 4. Visualización avanzada
  - Reportes dinámicos: *Rmarkdown* y *Sweave* (LaTeX).
  - Mapas
  - Gráficos especializados

### **Facilitadores:**

Dr. Luis Alberto Barboza Chinchilla.

Dr. Maikol Solís Chacón.

**Inversión:** \$450 por persona. Número mínimo de participantes: 8.

*Presupuesto:*