

PROGRAMA DE CURSO
Programación para Estadística I
XS0129

Características del curso

Horas:	4 (teoría-práctica)
Créditos:	4
Requisitos:	XS0124 Análisis Exploratorio de Datos XS0000 Lógica Para Programación y Demostraciones
Correquisitos:	Ninguno
Ciclo:	IV
Clasificación:	propio
Grado virtualidad:	Bajo virtual (máximo)
Profesor:	Michael Sánchez Soto (michael.sanchezsoto@ucr.ac.cr)
Horario clases:	Martes (virtual) 5:00pm – 7:00pm Jueves (presencial – LAB 140 CE) 5:00pm – 7:00pm
Consulta:	Lunes: 19:00 – 21:00 (Virtual)

Descripción

Este curso está dirigido a estudiantes de la carrera del Bachillerato en Estadística para que aprendan a programar en un lenguaje estadístico, así como el uso de control de versiones para el registro de cambios en el código y el trabajo colaborativo. El curso está orientado a adquirir destrezas en conceptos básicos de programación, manipulación de archivos de datos, realizar análisis básicos de estadística descriptiva y visualizarlos adecuadamente. Se trata de un curso que hace un uso intensivo del computador para habituar al estudiantado al ambiente de programación en R realizando trabajos prácticos grupales.

Objetivo general

Desarrollar destrezas en programación para manipular datos, realizar análisis básicos de estadística descriptiva y visualizarlos adecuadamente haciendo uso de control de versiones de código.

Objetivos específicos

Al finalizar el curso el/la estudiante estará en la capacidad de:

1. Utilizar software colaborativo para facilitar la programación en R.
2. Aplicar los conceptos básicos de programación para utilizarlos en el lenguaje R.
3. Aplicar funciones de las principales librerías de R en situaciones prácticas.

Habilidades y conocimientos (perfil de salida)

Habilidades	Conocimientos
HM01 - Manejar entidades matemáticas asociadas a la estadística	CM03 - Conocimientos básicos en lógica matemática
HE04 - Aprender de forma autónoma metodología estadística	CE08 - Conocimiento de estrategias de aprendizaje autodidacta
HT01 - Capturar, visualizar, procesar y analizar datos estructurados y no estructurados	CT04 - Conocimientos avanzados en manejo de herramientas de construcción de gráficos en capas (ej: ggplot2, plotly, matplotlib, seaborn)
HT02 - Crear, interpretar y modificar programas de código escrito en lenguajes de programación (ej: R, Python y SQL)	CT08 - Conocimientos intermedios en lenguajes de programación para análisis estadístico (ej: R) CT09 - Conocimientos básicos en diseño de algoritmos CT10 - Conocimientos básicos en Github como portafolio de proyectos para potenciales empleadores
HT03 - Crear y mantener archivos de datos y usar sistemas de bases de datos	CT14 - Conocimientos intermedios en limpieza y revisión de inconsistencias en archivos de datos

Contenidos

1. Aspectos básicos de la programación en R:
 - a) Conceptos básicos de lenguajes de programación.
 - b) Representación de algoritmos: diagramas de flujo.
 - c) R y RStudio, instalación y usos básicos.
 - d) Configuraciones globales de RStudio.
 - e) Uso básico de archivos de extensión .R y .Rmd y su relación con otros lenguajes de programación.

- f) Tipos de variables.
 - g) Operadores de asignación, aritméticos, lógicos, relacionales.
 - h) Estructuras de datos: vectores, matrices, arreglos, listas, conjuntos de datos.
 - i) Indexación y sustitución.
 - j) Estructuras de control iterativas (*for* y *while*).
 - k) Estructuras de control condicionales (*if*, *else*, *ifelse*).
 - l) Familia de funciones *apply*.
 - m) Diferencias entre mensajes, advertencias y errores.
2. Otros procedimientos para programar con R:
- a) Uso de librerías y funciones de R y el conflicto entre librerías.
 - b) Creación de funciones personalizadas en R.
 - c) Lectura y escritura de datos: *.csv*, *.xlsx*, *.sav*, *.txt*.
 - d) Uso de la librería *tidyverse*.
 - e) Procesamiento y manipulación de un *data.frame* con *dplyr*: selección, filtrado, agrupamiento, cálculo y modificación de variables.
 - f) Procesamiento y manipulación de un *data.frame* con *tidyr*: manipulación de formatos largo y ancho en un conjunto de datos.
 - g) Procesamiento y manipulación de datos temporales con *lubridate*.
 - h) Operadores de uso común: *%>%*, *%in%* y *%nin%*
 - i) Unión de datos con las funciones *left_join()*, *right_join()*, *inner_join()*, *anti_join()*, *full_join()* de la librería *dplyr* y con la función *do.call()*.
3. Visualización de datos:
- a) Construcción de gráficos con funciones base de R.
 - b) Construcción de gráficos por capas con *ggplot2*.
4. Análisis de datos:
- a) Medidas de posición y variabilidad en R.
 - b) Medidas de asociación y correlación en R.
 - c) Distribuciones de probabilidad en R: forma de sus funciones (*p_*, *r_*, *q_*).
 - d) Generación de variables aleatorias.
5. Versiones de control con Git:
- a) Control de cambios con Git.
 - b) Flujos de trabajo con Git.

- c) Colaboración con Git.
 - d) Diferencias entre Git y Github.
6. Productos de datos: Generación de reportes de extensión .html, .docx y .pdf.

Metodología

Este es un curso fundamentalmente práctico. Los contenidos del curso serán abordados en clase (laboratorio) y cada docente desarrollará actividades formativas para apoyar el proceso de enseñanza-aprendizaje. Para abordar los contenidos del curso el personal docente facilitará espacios para la puesta en práctica de los conocimientos adquiridos, por lo que se aplicará una metodología de un aprendizaje activo, donde cada estudiante debe realizar las actividades propuestas por su profesor o profesora, así como asistir y participar en los espacios de trabajo facilitados. El proceso para apoyar el aprendizaje del estudiantado incluye pruebas cortas y trabajos prácticos.

Se utilizará Mediación Virtual para ejecutar y evaluar las pruebas cortas, además, el material para la clase y material complementario se estarán subiendo a esta plataforma para disponibilidad del estudiante. También, se recurrirá a esta plataforma para la entrega de las tareas y exámenes que se ejecutaran durante el semestre, además, será el sitio donde podrán consultar el enlace de horas consulta y medio de comunicación activa.

Evaluación

La evaluación del curso se llevará a cabo mediante la aplicación de cuatro tipos de evaluación: quices, tareas programadas, un proyecto y exámenes. Cada una de las evaluaciones se desarrollan y calificaran de forma individual siguiendo los rubros indicados en las instrucciones dadas en cada una de las evaluaciones.

La distribución porcentual de las evaluaciones que se llevaran a cabo es la siguiente:

Evaluación	Porcentaje en la nota
Exámenes (3)	45%
Tareas programadas (2)	20%
Pruebas cortas (4)	20%

Proyecto	15%
----------	-----

Cronograma

Semana	Fechas	Contenido	Asignaciones / Comentario
1	10-12 marzo	Conceptos básicos de computación, algoritmos y lenguajes de programación	
2	17-19 marzo		
3	24-26 marzo	Introducción a R	Prueba corta 1
	31 marzo- 2 abril		Semana Santa
4	7-9 abril		Tarea 1 - Asignación
5	14-16 abril		Examen 1
6	21-23 abril		
7	28-30 abril	R Intermedio	Prueba corta 2
8	5-7 mayo		Tarea 1 – Entrega
9	12-14 mayo		
10	19-21 mayo		Examen 2
11	26-28 mayo	GIT	Tarea 2 - Asignación
12	2-4 junio	Análisis y Visualización de datos	Proyecto -Asignación
13	9-11 junio		Prueba corta 3
14	16-18 junio	Productos de datos	Tarea 2 - Entrega
15	23-25 junio		Prueba corta 4 Proyecto - Entrega

16	30 junio- 2 julio		Examen 3
	7 julio		Ampliación

Bibliografía

- Aguilar Fernández, E. & Zamora Araya, A. (2020). Introducción a la estadística descriptiva con R. <https://tinyurl.com/2oo9bmwk>
- Pace, L. (2012). Beginning R: An Introduction to Statistical Programming. <https://tinyurl.com/2qj7tyaz>
- Venables, W.N. & Smith, D. M. (2010). An Introduction to R <https://tinyurl.com/y9u5nxsy>
- Wickham, H. & Grolemund, G. (2018). R para Ciencia de Datos. <https://es.r4ds.hadley.nz/>