
PROGRAMA DE ESTUDIOS: CURSO DE PROGRAMACIÓN EN R

Área: Big Data

Programación en R

Información básica

[Información del curso]

1	Título	<i>Programación en R</i>
2	Año propuesto	<i>2024</i>
3	Semestre	<i>Primero (Agosto a Octubre)</i>
4	Departamento	<i>Centro de Innovación TIC (FPUNA)</i>
5	Nivel	<i>Básico</i>
6	Formato de clase	<i>Online</i>

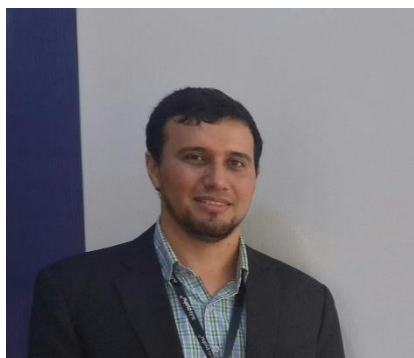
[Horario y Lugar]

1	Días	<i>Lunes y Jueves</i>
2	Horario	<i>de 17:00 a 19:30</i>
3	Ubicación	<i>Plataforma EDUCA</i>

[Información del instructor/a]

1	Nombre	<i>Prof. MSc. Juan Ignacio Mereles</i>
2	Oficina	<i>Departamento de Ciencias Básicas FP-UNA</i>
3	Contacto	<i>jmereles@pol.una.py</i>
4	Contacto	<i>(0985) 317-585</i>

[Perfil profesional]



Licenciado en Ciencias Mención Matemática Estadística y Licenciado en Ciencias Mención Matemática Pura, por la Facultad de Ciencias Exactas y Naturales (FACEN) de la UNA. Actualmente, en etapa de elaboración de tesis para obtener el título de Magíster en Estadística, también por la FACEN-UNA. Además, posee un magíster en Elaboración, Gestión y Evaluación de Proyectos de Investigación Científica. A lo largo de su carrera, ha participado y tutorizado varios cursos sobre análisis de datos con el uso de software estadístico y de programación de alto desempeño como lo son R y Python.

Cuenta con experiencia de más de ocho años en el campo de la Educación a Distancia, enseñando asignaturas en carreras de grado y posgrado, lo cual le ha permitido desarrollar habilidades pedagógicas sólidas y adaptarse a las necesidades de los estudiantes en entornos virtuales. Ha publicado artículos científicos en los campos de la educación y del análisis de datos, y es candidato a investigador por el CONACYT de Paraguay.

Prerrequisitos

Es preferible que los participantes tengan un conocimiento básico de estadística descriptiva. Aunque este aspecto es preferible, no afectará el desarrollo del curso si algún participante no cuenta con conocimientos de estadística descriptiva, pues durante el curso se desarrollará este tema con cierto detalle.

Descripción del curso

El curso "Programación en R" está diseñado para proporcionar a los estudiantes una base sólida en el uso del lenguaje de programación R para el análisis de datos. Este curso es adecuado tanto para principiantes sin experiencia previa en programación como para aquellos que buscan mejorar sus habilidades en el manejo de R. A lo largo del curso, se abordarán desde los conceptos básicos hasta técnicas avanzadas de manipulación y visualización de datos, así como análisis estadísticos. Las clases interactivas incluyen presentaciones teóricas y prácticas con sesiones de preguntas y respuestas en tiempo real, encuestas interactivas y estudios de casos. Los estudiantes participan activamente en discusiones y ejercicios prácticos, lo que les permite aplicar inmediatamente los conceptos aprendidos. El curso hará hincapié en la importancia de la ética en el análisis de datos. Los estudiantes aprenderán sobre la responsabilidad ética en la gestión y análisis de datos, incluyendo la privacidad y confidencialidad de la información, la transparencia en los procesos analíticos, y la obligación de reportar los resultados de manera honesta y precisa. Se discutirá la importancia de evitar el sesgo en el análisis de datos y la presentación de resultados, asegurando que las conclusiones sean objetivas y basadas en evidencia.

Objetivos del curso

Al finalizar con éxito este curso los estudiantes serán capaces de:

1. Comprender los conceptos básicos de programación en R, incluidos tipos de datos, variables y estructuras de control.
2. Manipular conjuntos de datos utilizando los paquetes dplyr y tidyr de R.
3. Realizar análisis estadísticos utilizando funciones y paquetes de R.

4. Crear y personalizar visualizaciones de datos usando ggplot2.
 5. Escribir y depurar scripts en R para diversas tareas de análisis de datos.
 6. Aplicar las mejores prácticas de programación en R y análisis de datos.
- Dominar los conceptos básicos de la programación en R: los participantes obtendrán una sólida comprensión del lenguaje de programación R, incluida su sintaxis, tipos de datos (vectores, matrices, listas, marcos de datos) y el entorno. Esta base es crucial para desarrollar habilidades más avanzadas en análisis y manipulación de datos.
 - Desarrollar competencia en manipulación y limpieza de datos: los estudiantes se volverán expertos en el uso de paquetes de R como dplyr y tidyr para una manipulación de datos eficiente. Esto incluye filtrar, ordenar, resumir y remodelar datos, así como manejar los valores faltantes y limpiar los datos para prepararlos para el análisis. El dominio de estas habilidades es esencial para transformar los datos sin procesar a un formato adecuado para el análisis.
 - Realizar análisis estadísticos: el curso equipará a los estudiantes con la capacidad de realizar análisis estadísticos descriptivos e inferenciales utilizando R. Esto incluye calcular medidas estadísticas (media, mediana, moda, desviación estándar), probar hipótesis y realizar análisis de regresión. Los estudiantes aprenderán a interpretar los resultados de pruebas estadísticas y a utilizarlos para tomar decisiones informadas basadas en datos.
 - Crear visualizaciones avanzadas de datos: los participantes aprenderán a usar ggplot2, un potente paquete de visualización en R, para crear visualizaciones informativas y estéticamente agradables. Esto incluye personalizar gráficos con temas, escalas y etiquetas, y crear una variedad de tipos de gráficos (gráficos de dispersión, histogramas, gráficos de barras, gráficos de líneas) para comunicar información valiosa sobre los datos de manera efectiva.
 - Aplicar conceptos de programación en R: los estudiantes explorarán conceptos de programación más avanzados en R, como escritura de funciones, estructuras de control (declaraciones if-else, bucles) y manejo de errores. Estas habilidades son importantes para escribir scripts R eficientes que automaticen las tareas de análisis de datos y mejoren la reproducibilidad.
 - Trabajar con tipos de datos avanzados: el curso introducirá el manejo de estructuras y tipos de datos más complejos, como datos de series de tiempo, datos de texto y datos espaciales. Los estudiantes aprenderán métodos y paquetes apropiados para analizar este tipo de datos, ampliando la gama de análisis que pueden realizar.
 - Implementar mejores prácticas de programación en R: a lo largo del curso, los estudiantes conocerán las mejores prácticas en codificación, análisis de datos y visualización. Esto incluye escribir código limpio y legible, documentar el trabajo para su reproducibilidad y consideraciones éticas en el análisis de datos.

- Completar un proyecto final: al final del curso, los participantes aplicarán las habilidades y conocimientos adquiridos a un proyecto de análisis de datos del mundo real. Este proyecto final requerirá que los estudiantes limpien, analicen y visualicen un conjunto de datos y presenten sus hallazgos, demostrando su capacidad para llevar a cabo un flujo de trabajo de análisis de datos integral utilizando R.

Método de evaluación

La calificación del curso está estructurada para evaluar la comprensión, el compromiso y la aplicación práctica de los materiales del curso por parte de los estudiantes. La política de calificaciones incluye:

- Asistencia: 10%
- Tareas y cuestionarios: 30%
- Examen parcial: 20%
- Examen final: 30%
- Participación: 10%

Libros de texto y otros materiales necesarios

- Chang, W. (2018). *R graphics cookbook: practical recipes for visualizing data*. O'Reilly Media. <https://r-graphics.org/>
- Hernández, F., & Usuga, O. (2024). Manual de R. <https://fhernanb.github.io/Manual-de-R/>
- Hernández, F., Usuga, O., & Mazo, M. (2024). Modelos de Regresión con R. https://fhernanb.github.io/libro_regresion/
- Irizarry, R. A. *Introduction to Data Science: Data Analysis and Prediction. Algorithms with R*. CRC Press. <https://rafalab.dfci.harvard.edu/dsbook-part-1/> and <https://rafalab.dfci.harvard.edu/dsbook-part-2/>
- Ismay, C., & Kim, A. Y. (2019). *Statistical inference via data science: a ModernDive into R and the tidyverse*. Chapman and Hall/CRC. <https://moderndive.com/index.html>
- Santana, S., & Farfán, E. (2014). *El arte de programar en R: un lenguaje para la estadística*. https://cran.r-project.org/doc/contrib/Santana_El_arte_de_programar_en_R.pdf
- Wickham, H., Çetinkaya-Rundel, M., & Grolemund, G. (2023). *R for Data Science*. O'Reilly Media, Inc. <https://r4ds.hadley.nz/>

Tarea(s) y examen(es)

Las tareas y exámenes son componentes esenciales del curso "Programación en R" y están diseñados para reforzar los conceptos teóricos tratados en las clases, así como para proporcionar experiencia práctica en el manejo de R para análisis de datos. A continuación, se detallan los tipos de tareas y los componentes de los exámenes, junto con una descripción detallada de cada uno.

Tareas de programación: Los estudiantes recibirán tareas semanales que implican la escritura de scripts en R para realizar diversas operaciones, como manipulación de datos, análisis estadísticos y creación de visualizaciones. Éstas están diseñadas para desarrollar habilidades prácticas y fortalecer la comprensión de los conceptos aprendidos en clase.

Proyectos prácticos: A lo largo del curso, los estudiantes trabajarán en proyectos más grandes que integran múltiples conceptos y técnicas. Estos proyectos pueden incluir el análisis de un conjunto de datos real, donde los estudiantes deben aplicar técnicas de manipulación de datos, análisis estadísticos y visualización.

Cuestionarios en línea: Los estudiantes completarán cuestionarios en línea que evalúan su comprensión de los conceptos teóricos. Estos cuestionarios pueden incluir preguntas de opción múltiple, verdadero/falso y preguntas de desarrollo corto.

Examen parcial: Se realizará a mitad del curso para evaluar la comprensión de los conceptos básicos de programación en R, manipulación de datos y visualización. El examen incluirá preguntas teóricas y prácticas donde los estudiantes deberán escribir y depurar scripts en R.

Examen final: Al finalizar el curso, se realizará un examen comprensivo que cubre todos los temas tratados, incluyendo análisis estadísticos y técnicas avanzadas de programación. Este examen evaluará la capacidad de los estudiantes para integrar y aplicar los conocimientos adquiridos a lo largo del curso.

Proyecto Final: Consiste en la realización de un trabajo sobre los contenidos desarrollados en el módulo, utilizando herramientas de generación de documentos dinámicos integradas en RStudio: RMarkdown o Quarto.

El docente del curso proporcionará las indicaciones para realizar el trabajo en la segunda semana, una vez que los estudiantes hayan reconocido los aspectos básicos del software. Estas indicaciones incluirán el periodo de entrega, que será hasta dos semanas después de la culminación de las clases sincrónicas, el medio de entrega y los temas a tratar en el trabajo.

Actividades del curso

Las actividades del curso están diseñadas para que los estudiantes se involucren activamente con los materiales, refuercen su comprensión y desarrollen habilidades prácticas. Estas actividades van desde conferencias y debates interactivos hasta laboratorios prácticos y proyectos en grupo.

- Las clases interactivas incluyen, entre otras cosas, sesiones de preguntas y respuestas y encuestas en tiempo real.
- Las sesiones de debate incluyen, entre otras cosas, estudios de casos y discusiones.
- Sesiones prácticas: ejercicios de codificación, desafíos de datos, etc.
- Los proyectos en grupo incluyen, entre otras cosas, talleres sobre proyectos y revisiones por pares (peer review).
- Los talleres sobre temas especializados pueden incluir tutoriales sobre herramientas específicas, etc.
- Las actividades de aprendizaje entre pares incluyen, entre otras, grupos de estudio, clubes de codificación, etc.

Cada una de estas actividades está diseñada para complementar los conocimientos teóricos adquiridos en las clases, profundizar en la comprensión mediante el debate y la aplicación, y preparar a los estudiantes para tareas relacionadas con el mundo real. Mediante la participación en diversas actividades del curso, los estudiantes desarrollan un conjunto completo de habilidades que incluyen la competencia técnica, el razonamiento ético y las experiencias de trabajo en colaboración.

Cronograma del curso

Semana	Tema	Tipo de clases	Materiales
1	Tipos y estructuras de datos	<i>Clase interactiva</i>	<ul style="list-style-type: none"> ● Introduction to Data Science: Data Analysis and Prediction. Algorithms with R ● El arte de programar en R: un lenguaje para la estadística
2	Manipulación de datos	<i>Clase interactiva</i>	<ul style="list-style-type: none"> ● Introduction to Data Science: Data Analysis and Prediction. Algorithms with R ● R for Data Science
3	Estadística básica con R	<i>Debate</i>	<ul style="list-style-type: none"> ● Manual de R
4	Visualización de datos	<i>Laboratorios prácticos</i> <i>Examen parcial</i>	<ul style="list-style-type: none"> ● R for Data Science ● R graphics cookbook: practical recipes for visualizing data
5	Manipulación avanzada de datos	<i>Clase interactiva</i>	<ul style="list-style-type: none"> ● Introduction to Data Science: Data Analysis and Prediction. Algorithms with R ● R for Data Science
6	Conceptos de programación	<i>Clase interactiva</i>	<ul style="list-style-type: none"> ● El arte de programar en R: un lenguaje para la estadística ● R for Data Science
7	Contrastes de hipótesis estadísticas	<i>Clase interactiva</i>	<ul style="list-style-type: none"> ● Statistical inference via data science: a ModernDive into R and the tidyverse
8	Modelación estadística	<i>Taller</i> <i>Examen final</i>	<ul style="list-style-type: none"> ● Introduction to Data Science: Data Analysis and Prediction. Algorithms with R ● Modelos de Regresión con R

* Un curso está diseñado para un periodo de 8 semanas y un total de 40 horas requeridas (5 horas por semana).

Contenidos del curso

Semana 1: Tipos y estructuras de datos

Objetivo: Introducir a los estudiantes en los tipos y estructuras de datos en R, comprendiendo su importancia y uso en el análisis de datos.

Tipo de clase: Clase interactiva

Actividades:

- Introducción a los tipos de datos básicos en R: numéricos, caracteres, lógicos, factores.
- Trabajo con vectores, listas, matrices y data frames.
- Ejercicios prácticos para identificar y manipular diferentes tipos de datos.

Materiales:

- Introduction to Data Science: Data Analysis and Prediction Algorithms with R: Capítulo 2
- El arte de programar en R: un lenguaje para la estadística: Capítulo 2.

Semana 2: Manipulación de datos

Objetivo: Aprender técnicas esenciales para la manipulación y transformación de datos en R.

Tipo de clase: Clase interactiva

Actividades:

- Uso de dplyr para la manipulación de data frames.
- Funciones select(), filter(), mutate(), arrange(), summarise().
- Combinación de múltiples funciones usando el operador pipe (%>% o |>).

Materiales:

- Introduction to Data Science: Data Analysis and Prediction Algorithms with R: Capítulo 4.
- R for Data Science: Capítulo 3.

Semana 3: Estadística básica con R

Objetivo: Introducir conceptos básicos de estadística y su implementación en R.

Tipo de clase: Debate

Actividades:

- Descripción de medidas de tendencia central y dispersión.
- Cálculo de estadísticas descriptivas.
- Discusión sobre la interpretación de resultados estadísticos.

Materiales:

- Manual de R: Capítulos 9 al 14

Semana 4: Visualización de datos

Objetivo: Desarrollar habilidades para la visualización de datos usando ggplot2 en R.

Tipo de clase: Laboratorios prácticos

Actividades:

- Introducción a ggplot2 y su gramática de gráficos.
- Creación de gráficos básicos: histogramas, scatter plots, box plots.
- Personalización de gráficos: temas, etiquetas, escalas.

Materiales:

- R for Data Science: Capítulos 1, 9 y 10.
- R graphics cookbook: practical recipes for visualizing data: Capítulos 3 al 14.

Examen parcial: Evaluación de conceptos aprendidos en las primeras cuatro semanas.

Semana 5: Manipulación avanzada de datos

Objetivo: Profundizar en técnicas avanzadas de manipulación de datos en R.

Tipo de clase: Clase interactiva

Actividades:

- Trabajo con funciones avanzadas de dplyr y tidyr.
- Manipulación y limpieza de datos faltantes.
- Reshape de data frames utilizando gather() y spread().

Materiales:

- Introduction to Data Science: Data Analysis and Prediction Algorithms with R: Capítulo 4.
- R for Data Science: Capítulos 3 y 5.

Semana 6: Conceptos de programación

Objetivo: Enseñar conceptos fundamentales de programación en R para la creación de scripts y funciones.

Tipo de clase: Clase interactiva

Actividades:

- Introducción a la programación con R: estructuras de control, bucles y condicionales.
- Creación y uso de funciones personalizadas.
- Buenas prácticas de programación.

Materiales:

- El arte de programar en R: un lenguaje para la estadística: Capítulos 4 y 5.
- R for Data Science: Capítulos 25 y 26.

Semana 7: Contrastes de hipótesis estadísticas

Objetivo: Introducir a los estudiantes en los métodos de inferencia estadística y contraste de hipótesis.

Tipo de clase: Clase interactiva

Actividades:

- Explicación de conceptos de hipótesis nula y alternativa.
- Pruebas t, ANOVA y chi-cuadrado.
- Interpretación de resultados y p-valores.

Materiales:

- Statistical inference via data science: a ModernDive into R and the tidyverse: Capítulo 9.

Semana 8: Modelación estadística

Objetivo: Introducir conceptos de modelación estadística y su aplicación en R.

Tipo de clase: Taller

Actividades:

- Regresión lineal simple y múltiple.
- Evaluación del modelo: R-cuadrado, errores residuales.
- Interpretación de coeficientes y predicciones.

Materiales:

- Introduction to Data Science: Data Analysis and Prediction Algorithms with R: Capítulo 14.
- Modelos de Regresión con R: Capítulos 1 al 6.

Examen final: Evaluación de todos los conceptos aprendidos a lo largo del curso.