



CENTRO MICRODATOS  
Departamento de Economía

## CURSO ANÁLISIS DE DATOS CON R NIVEL INTERMEDIO

### 1. Descripción del Curso

Este curso busca profundizar la práctica y manejo del software R, poniendo especial énfasis en la creación de documentos finales y la automatización de procesos. Entre los tópicos a discutir se encuentran: uso de iteraciones en de manera avanzada para la programación, generación de funciones complejas, visualización gráfica avanzada, generación automática de documentos Excel, html y presentaciones. Se pondrá especial énfasis en la enseñanza de la programación, con el fin de generar soluciones en la automatización y eficiencia de procesos estadísticos utilizando como base encuestas y bases de datos con gran nivel de información.

### 2. Objetivos

1. Realizar un análisis exploratorio de los datos, usando gráficos, mapas y elementos de estadística descriptiva de manera avanzada
2. Ser capaces de utilizar loops y todas las funciones asociadas
3. Crear documentos Excel desde R
4. Generación de informes, presentaciones y documentos presentables sofisticados.

### 3. Modalidad

Sincrónica. Las clases se realizarán mediante la plataforma Webex

### 4. Duración

24 horas, distribuidas en 6 días.

## 5. Contenidos

### **Clase 1:** Recordatorio de R (duración: 4 horas)

- Manipulación de bases de datos (wide y long data)
- Generación de variables (continuas, binarias y discretas)
- Variables de fechas
- Agrupación, recodificación, transformaciones, uso paquete dplyr
- Generación de variables mediante iteraciones (Loops)
- Control de iteraciones (if-else; for; while; break)

### **Clase 2:** Automatización (duración: 4 horas)

- Generación de funciones
- Uso de iteraciones dentro de funciones
- Crear gráficos mediante loops
- Generación y división de bases de datos mediante loops

### **Clase 3:** Estadística descriptiva (duración: 4 horas)

- Estimaciones de medias, varianzas, intervalos de confianza, etc.
- Gráfico de puntos, lollipop, densidad y barras, uso de paquete ggalt
- Gráficos de calor, uso de mapas.
- Manipulación de datos para generación de estadística descriptiva presentada en gráficos
- Uso de factores de expansión

### **Clase 4** Manipulación de datos (duración: 4 horas)

- Generación de nubes de palabras, text mining
- Manipulación de datos para generar tablas exportables avanzadas
- Transformaciones estadísticas

### **Clase 5:** Excel (duración: 4 horas)

- Web scrapping como ejemplo de recolección de datos (con APIs)
- Exportar dataframes en formato xlsx, uso de paquete openxlsx
- Manipulación de Excel desde R para generar documentos con índice, tablas, títulos, etc.

### **Clase 6:** RMarkdown (duración: 4 horas)

- Web scrapping manual con ejemplo Banco Central (sin APIs)
- Creación de documentos html
- Creación de presentaciones mediante RMarkdown, uso de Shiny

## 6. Requisitos del curso

Realización de curso introductorio de R:

- Saber manipular bases de datos,
- Importar y exportar datos,
- Conocimiento de paquete dplyr, stringr, ggplot, tidyr, magrittr

## 7. Proyecto de clase

Durante el transcurso del curso, se espera que las y los estudiantes realicen un proyecto individual, el cual tiene como objetivo replicar boletines de información como el siguiente:

[https://ine.cl/docs/default-source/ocupacion-y-desocupacion/boletines/2021/pa%C3%ADs/ene-nacional-son21-vf\\_2021-12-29.pdf?sfvrsn=3352f6f2\\_4](https://ine.cl/docs/default-source/ocupacion-y-desocupacion/boletines/2021/pa%C3%ADs/ene-nacional-son21-vf_2021-12-29.pdf?sfvrsn=3352f6f2_4)

Para ello, deberán escoger una base de datos y realizar las estimaciones de la manera más automática posible. Los detalles de la información necesaria para el proyecto se entregarán en la primera clase del curso.

## 8. Certificación del curso

El curso requiere de un 75% de asistencia y la aprobación del proyecto de clase.