
Curso Académico: (2023 / 2024)**Fecha de revisión: 15-07-2023**

Departamento asignado a la asignatura: Departamento de Estadística**Coordinador/a: UCAR MARQUES, IÑAKI****Tipo: Obligatoria Créditos ECTS : 6.0****Curso : 1 Cuatrimestre : 1**

REQUISITOS (ASIGNATURAS O MATERIAS CUYO CONOCIMIENTO SE PRESUPONE)

Introducción a la Programación con R (19151)

OBJETIVOS

- Conocimiento de las estructuras y procedimientos propios de la programación con datos.
- Capacidad de importar datos tabulares en una variedad de formatos con el lenguaje de programación R.
- Capacidad de trabajar con bases de datos remotas.
- Capacidad de preparar, limpiar, transformar y enriquecer datos tabulares para su posterior modelado y visualización con los lenguajes de programación R y SQL.

DESCRIPCIÓN DE CONTENIDOS: PROGRAMA

1. Programación en base R
 - 1.1. Introducción al ecosistema R
 - 1.2. Introducción a RStudio. Trabajando con proyectos
 - 1.3. Tipos de datos básicos
 - 1.4. Primeros usos de funciones y paquetes. Operaciones básicas
 - 1.5. Entendiendo los errores
2. De la celda al conjunto de datos
 - 2.1. Concatenar valores: vectores (variables)
 - 2.2. Operaciones básicas con vectores
 - 2.3. Bucles vs. programación vectorial. Estructuras de control de flujo
 - 2.4. Primeras bases de datos: matrices y data.frames
 - 2.5. Tibbles como tipo estándar de base de datos. El paquete datapasta
3. Tidy data
 - 3.1. R base vs. tidyverse. El operador pipe
 - 3.2. Principios de tidy data: tidy vs. messy data
 - 3.3. Pivotando conjuntos de datos
4. RMarkdown y quarto: reportar resultados
5. Tidyverse
 - 5.1. Operaciones por filas. Limpiando datos: valores perdidos y duplicados
 - 5.2. Operaciones por columnas
 - 5.3. Agregaciones y recategorización de variables
 - 5.4. Agrupaciones: group_by y .by
 - 5.5. Sumarios
 - 5.6. Uniendo conjuntos de datos
 - 5.7. Importación/exportación desde/a diferentes formatos
 - 5.8. Uso de APIs
6. Tipos de datos avanzados
 - 6.1. Variables categóricas: paquete forcats
 - 6.2. Tratamiento de cadenas: paquete stringr
 - 6.3. Tratamiento de fechas: paquete lubridate
 - 6.4. Tratamiento de listas: paquete purrr. Programación funcional
7. Gestión avanzada de datos

- 7.1. Paquete dbplyr: del tidyverse a SQL
- 7.2. Paquete arrow: tratamiento de bases de datos masivas

- 8. Programación en SQL
 - 8.1. Introducción a las bases de datos relacionales
 - 8.2. Manipulación de datos y consultas
 - 8.3. Consultas complejas, agregaciones y subconsultas
 - 8.4. Combinaciones de múltiples tablas

ACTIVIDADES FORMATIVAS, METODOLOGÍA A UTILIZAR Y RÉGIMEN DE TUTORÍAS

Actividades Formativas:

- Clases teórico-prácticas
- Tutorías
- Trabajo en grupo
- Trabajo individual del estudiante

Metodologías Docentes:

- Exposiciones en clase del profesor con soporte de medios informáticos y audiovisuales, en las que se desarrollan los conceptos principales de la materia y se proporciona la bibliografía para complementar el aprendizaje de los alumnos.
- Resolución de casos prácticos, problemas, etc., planteados por el profesor de manera individual o en grupo.
- Exposición y discusión en clase, bajo la moderación del profesor de temas relacionados con el contenido de la materia, así como de casos prácticos.
- Elaboración de trabajos e informes de manera individual o en grupo.

SISTEMA DE EVALUACIÓN

- Participación en clase (10%)
- Trabajos individuales realizados durante el curso (60%)
- Trabajo en grupo realizado al final del curso (30%)

En la convocatoria extraordinaria, el sistema de evaluación será el siguiente:

1) Examen: 100%

Peso porcentual del Examen Final:	0
Peso porcentual del resto de la evaluación:	100

BIBLIOGRAFÍA BÁSICA

- Hadley Wickham R for Data Science, O'Reilly, 2017

BIBLIOGRAFÍA COMPLEMENTARIA

- Chester Ismay and Albert Y. Kim Statistical Inference via Data Science: a Modern Dive into R and the tidyverse, Chapman & Hall, 2022
- Steph Locke Data Manipulation in R, Locke Data, 2017