

PROGRAMA - AÑO 2018			
Espacio Curricular:	Análisis de Datos Científicos en R		
Carácter:	Electivo	Período d	1º semestre
Carrera/s:	Cursos de Posgrado: especialización, maestría o doctorado en Ciencia y Tecnología, Biología, Física, Matemática, Geología, Química, y/o áreas relacionadas.		
Profesor Responsable:	Luis Gregorio Moyano		
Equipo Docente:			
Carga Horaria: 120 hs - 32 hs teóricas presenciales y 88 hs prácticas (16 presenciales y 72 horas no presenciales).			
Requisitos de Cursado:	Materias recomendadas: Probabilidad y Estadística, Programación en C.		

### 1-EXPECTATIVAS DE LOGRO

- Dominar los conceptos básicos para el manejo del lenguaje R.
- Desarrollar e implementar análisis de datos empíricos, mediante scripts, en el lenguaje R.
- Comprender las ventajas y desventajas del lenguaje R y como se complementa con otras técnicas de análisis de datos.
- Conocer las fuentes de información y datos necesarias para acceder de manera independiente al conocimiento requerido para resolver problemas científicos específicos.
- Desarrollar e implementar modelos científicos para resolver problemas simples en diferentes áreas del conocimiento científico.
- Demostrar la utilización de técnicas avanzadas de programación en R para el modelado y el análisis de datos.

### 2-DESCRIPTORES

El rol de R en el análisis de datos. Estructuras de datos y controles de flujo. Funciones y paquetes. Vectorización. Complejidad computacional y eficiencia. Paralelización. Análisis de datos. Visualización con aplicaciones científicas. Conexión con bases de datos. Integración con C. Paquetes científicos especializados.

### **3-CONTENIDOS ANALÍTICOS**

#### **Unidad 1: Introducción general**

Introducción y fundamentos. El rol de R en el análisis de datos. Ventajas y limitaciones. Lenguajes de *scripting*. Flujos de trabajo científicos, git y github. Análisis interactivo de datos científicos.

#### **Unidad 2: Tipos de datos y control de flujo**

Estructuras de datos fundamentales. *Subsetting*. Operadores. Controles de flujo.

#### **Unidad 3: Funciones y gestión paquetes externos**

Funciones. Gestión de paquetes. Programación funcional. Vectorización y complejidad computacional. Paquete Rcpp e integración con C y C++. *Hadleyverse*: dplyr, purrr, stringr, lubridate, tydyr, devtools. *Piping* con magrittr.

#### **Unidad 4: Visualización de datos: ggplot2**

Introducción al paquete ggplot2. Gramática de visualización. /Faceting/. Conexión con bases de datos: RMySQL, RSQLite, rvest. Elementos de Shiny para interacción con datos en tiempo real.

#### **Unidad 5: Análisis de datos científicos - Aplicaciones**

Ejemplos prácticos de datos empíricos. Paquetes de datos, fuentes de datos públicos y métodos de acceso online. Paquetes especializados para Biología, Física, Matemática, Geología y Química. Análisis de redes con *igraph* con aplicaciones. Rol de R en Estadística. Rol de R en Bioinformática.

#### **Unidad 6: Preparación de trabajo final.**

Elección de un tema de investigación, en lo posible alineado con los intereses y/o necesidades del alumno o de la alumna. Planteo del problema y búsqueda de paquetes disponibles específicos para su resolución. Desarrollo del trabajo final en formato de publicación científica. Preparación de seminario final con nivel de curso de posgrado.

#### 4-BIBLIOGRAFÍA

1. Wickham, Hadley. Advanced R. CRC Press, 2014.
2. Gerrard, Paul, and Radia M. Johnson. Mastering scientific computing with R. Packt Publishing Ltd, 2015.
3. Gentleman, Robert. R programming for bioinformatics. CRC Press, 2008.
4. Van der Loo, Mark PJ. Learning RStudio for R statistical computing. Packt Publishing Ltd, 2012.
5. Wickham, Hadley. "ggplot2." Wiley Interdisciplinary Reviews: Computational Statistics 3.2 (2011): 180-185.
6. Ginestet, Cedric. "ggplot2: elegant graphics for data analysis." Journal of the Royal Statistical Society: Series A (Statistics in Society) 174.1 (2011): 245-246.
7. Alanis-Lobato, Gregorio, Pablo Mier, and Miguel A. Andrade-Navarro. "Efficient embedding of complex networks to hyperbolic space via their Laplacian." Scientific Reports 6 (2016).
8. Kabacoff, Robert. R in action: data analysis and graphics with R. Manning Publications Co., 2015.
9. Chang, Winston. R graphics cookbook. "O'Reilly Media, Inc.", 2012.
10. Zumel, Nina, John Mount, and Jim Porzak. Practical data science with R. Manning, 2014.
11. Wickham, Hadley. "Making an R package." (2010).
12. Chacon, Scott, and Ben Straub. Pro GIT. Apress, 2014.
13. Hunt, Andrew, and David Thomas. The pragmatic programmer: from journeyman to master. Addison-Wesley Professional, 2000.
14. Wickham, Hadley. "The split-apply-combine strategy for data analysis." Journal of Statistical Software 40.1 (2011): 1-29.

#### 5-METODOLOGÍA DE ENSEÑANZA Y EVALUACIÓN DURANTE EL CURSADO

El contenido de la materia será introducido mediante presentaciones orales a cargo del docente. Además se realizarán clases prácticas en la sala de computación, donde las alumnas y alumnos aplicarán los conocimientos adquiridos resolviendo guías de problemas. La evaluación de la alumna o alumno será de la siguiente manera:

- Dos evaluaciones parciales sobre conocimientos generales de la materia (una nota cada uno).
- Elaboración, desarrollo y presentación de un trabajo final integrador (una nota).

Las tres notas deberán ser mayores a 6 (seis). La nota final será un promedio ponderado de las tres notas anteriores.

## 6- CONDICIONES DE REGULARIDAD TRAS EL CURSADO

Son requisitos para que alumnas y alumnos sean considerados regulares: aprobar las evaluaciones parciales y el trabajo final.

## 7- SISTEMA DE APROBACIÓN Y/O PROMOCIÓN DEL ESPACIO CURRICULAR

La asignatura se considerará promocionada cuando se aprueben con una nota igual o superior a 6 (seis) tanto las evaluaciones escritas como la presentación del trabajo final. El trabajo final de investigación se enfocará en la utilización de técnicas avanzadas de programación en R, por lo que se requerirá un desarrollo mayor y el uso de técnicas más complejas de programación para la elaboración de dicho trabajo final. El tema del trabajo final podrá ser alineado con los temas de investigación del alumno de posgrado, por ejemplo, vinculado a su tesis.

### MODALIDAD DE EXAMEN PARA ALUMNOS LIBRES

- Aprobar un examen escrito sobre problemas propios de la asignatura, cubriendo la totalidad de la misma.
- Aprobar un examen oral sobre la teoría dictada regularmente y de acuerdo a los contenidos analíticos del programa vigente.
- Presentar un proyecto equivalente al trabajo final de la materia (tanto informe final como exposición oral con un seminario).

Las notas obtenidas en cada uno de los tres puntos detallados anteriormente deberá ser igual o superior a 6 (seis). La nota final será un promedio ponderado de las tres calificaciones obtenidas.

**PROMOCIONABLE** *(Marque con una cruz la respuesta correcta)*

SI

X

NO



FIRMA Y ACLARACIÓN

DEL RESPONSABLE DEL ESPACIO CURRICULAR