

PROGRAMA - AÑO 2019			
Espacio Curricular:	Análisis de Datos Científicos en R		
Carácter:	Electivo	Período	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:	Sede Central: Luis Gregorio Moyano Docente invitada: Carolina Gandini (Beca Doctoral CONICET, IBAM-CONICET)		
Carga Horaria: 120 hs - 36 hs presenciales y 84 hs no presenciales.			
Requisitos de Cursado:	Materias recomendadas: Probabilidad y Estadística, Programación en Python.		

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. *Tidyverse*. *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.

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-COMPETENCIAS		
Tipo de Competencias	Detalle	Articula con:
Genéricas	<ol style="list-style-type: none"> 1. Manejo de herramientas informáticas para el análisis y visualización de datos. 2. Resolución numérica de problemas concretos en biología, física y química mediante herramientas informáticas. 	<ol style="list-style-type: none"> 1. Introducción a la programación en Python. 2. Probabilidad y Estadística, Elementos de Cálculo Numérico, Aplicaciones de Ecuaciones Diferenciales. 3. Bioestadística, Bioinformática, 4. Física Computacional, Simulaciones numéricas en las ciencias básicas Física Experimental I y II.
Específicas	<ol style="list-style-type: none"> 1. Comprensión de la lógica de un programa de computación. 	<ol style="list-style-type: none"> 1. Introducción a la programación en Python, Física Computacional, Simulaciones numéricas en las ciencias básicas, Elementos de Cálculo Numérico.

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. Las clases se desarrollarán en el laboratorio de informática. Cada estudiante podrá utilizar su computador personal, para lo cuales se le dará el soporte para la instalación de las herramientas informáticas necesarias. Se realizarán clases prácticas donde cada estudiante aplicará los conocimientos adquiridos resolviendo guías de problemas. La evaluación será de la siguiente manera:

- Dos evaluaciones parciales, obligatorias para la promoción, sobre conocimientos dados hasta el momento. La primera evaluación será sin nota a mitad del cursado, con el objetivo de proveer a la alumna o alumno de una referencia sobre el nivel alcanzado hasta ese momento, y evaluar estrategias para la segunda parte del curso. Esta evaluación no afectará la nota de la evaluación final. El segundo parcial será con nota.
- Elaboración, desarrollo y presentación de un trabajo final integrador escrito, con nota.

El trabajo final de integrador 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, siendo necesario una mayor cantidad de horas de dedicación de los estudiantes de posgrado fuera del curso. El tema del trabajo final podrá ser alineado con los temas de investigación del alumno de posgrado, por ejemplo, vinculado a su tesis.

Se recibirán pedidos de consulta durante la duración del semestre.

6- CONDICIONES DE REGULARIDAD TRAS EL CURSADO

Condiciones de **regularidad**:

- Asistencia a los exámenes parciales. Aprobar el trabajo final.

Condiciones de **promoción**:

- Aprobar el segundo examen parcial con nota mayor o igual a 6 (seis). Aprobar el trabajo final con nota mayor o igual a 6 (seis).



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 la evaluación escrita (2do. parcial) como el trabajo final.

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 de carácter intensivo sobre el material teórico dictado durante todo el curso y de acuerdo a los contenidos analíticos del programa vigente.
- Presentar un proyecto equivalente al trabajo final de la materia, tanto informe escrito como exposición oral de tipo 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 <i>(Marque con una cruz la respuesta correcta)</i>	SI	X	NO	
---	----	---	----	--

FIRMA Y ACLARACIÓN

DEL RESPONSABLE DEL ESPACIO CURRICULAR