

Aspectos generales

Título:	Análisis de datos para estudios biomédicos con R
Programas de posgrado o planes de estudio en donde se ofertará adicionalmente:	Ninguno
Área del conocimiento:	Biología celular
Semestre:	2024-2
Modalidad:	Tópico selecto
Horario:	Jueves 4-7 pm
No. sesiones:	16
Horas por sesión:	3.0
Total alumnos PDCB:	20
Total alumnos:	20
Videoconferencia:	No
Lugar donde se imparte:	Unidad de posgrado
Informes:	federico.castro@st.ib.unam.mx

Métodos de evaluación

MÉTODO	PORCENTAJE	NOTAS
Exámenes	30%	Dos
Presentación de un proyecto	10%	
Trabajos	60%	Dos

Contribución de este curso/tópico en la formación del alumnado del PDCB:

El curso está dirigido a estudiantes de posgrado en ciencias biomédicas y busca proveer al alumno con las herramientas necesarias para que pueda estructurar y trabajar con datos que utilizará en su tesis o en su labor profesional. El curso tiene una componente práctico fuerte en el que el estudiante puede trabajar con datos propios o con datos de su elección bajo la asesoría de los investigadores que imparten esta materia.

Profesor (a) responsable

Nombre:	Castro Monzon Federico
Teléfono:	5510772464
Email:	federico.castro@st.ib.unam.mx

Profesores (as) participantes

PARTICIPANTE	ENTIDAD O ADSCRIPCIÓN	SESIONES
CASTRO MONZON FEDERICO Responsable	Otras entidades	2.6 Repaso de subselección 2.3 Manejo de datos múltiples 8.1 Trabajo con asesoría 1.0 Introducción a R 1.6 Cómo obtener ayuda 2.0 Manipulación básica de datos 3.0 Importar y exportar datos 4.0 Análisis exploratorio de datos

SÁNCHEZ DOMÍNGUEZ MARIO
Integrante

Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas / Facultad de Ciencias, UNAM

8.0 Desarrollo del proyecto personal
4.4 Visualización de datos
6.0 Modelos
7.0 Temas avanzados y repaso
8.1 Trabajo con asesoría
5.0 Programación
5.2 Funciones personalizadas
6.3 Modelos multifactoriales

Introducción

En este curso tiene como objetivo ofrecer las herramientas básicas para el uso de R, un lenguaje libre de programación usado frecuentemente en la academia para la exploración, visualización y análisis de datos. El programa cuenta con el soporte de una amplia comunidad y gran cantidad de librerías han sido desarrolladas para el análisis de datos en campos como epidemiología, ecología, fisiología y genética.

Temario

Clase 1, Federico Castro Monzón

- 1.0 Introducción a R
- 1.1 Historia de R
- 1.2 Filosofía del lenguaje
- 1.3 Ejemplos de uso en estudios de biomedicos
- 1.4 Instalación de R en Windows, Mac y Linux
- 1.5 Entornos de desarrollo integrado

Clase 2, Federico Castro Monzón

- 1.6 Cómo obtener ayuda
- 1.6.1 Ayuda en R
- 1.6.2 StackOverflow
- 1.6.3 Librería Swirl
- 1.7 Propuesta de proyecto personal

Clase 3, Federico Castro Monzón

- 2.0 Manipulación básica de datos
- 2.1 Tipos de datos
- 2.1.1 Elementos numéricos
- 2.1.2 Elementos booleanos
- 2.1.3 Caracteres
- 2.1.4 Factores
- 2.2 Operaciones básicas
- 1.7 Propuesta de proyecto personal

Clase 4, Federico Castro Monzón

- 2.3 Manejo de datos múltiples
- 2.3.1 Vectores
- 2.3.2 Matrices
- 2.3.3 Tablas
- 2.3.4 Listas
- 2.4 Coerción de datos
- 2.5 Sub-selección de datos
- 1.7 Propuesta de proyecto personal

Clase 5, Federico Castro Monzón

- 2.6 Repaso de subselección

Clase 6, Federico

- 3.0 Importar y exportar datos
- 3.1 Tipos de formatos de tablas
- 3.2 Conversión a Excel/Libre Office
- 3.3 Establecer el directorio base
- 3.4 Importar tablas
- 3.5 Errores comunes en archivos
- 3.6 Exportar tablas

Clase 7, Federico Castro Monzón

- 4.0 Análisis exploratorio de datos

4.1 Detección de datos ausentes (NA)

4.3 Estadísticas básicas

4.3.1 Dispersión

4.3.2 Centralización

4.3.3 Correlación

Clase 8, Mario Sánchez Domínguez

4.4 Visualización de datos

4.4.1 Diagramas de dispersión

4.4.2 Diagramas de caja

4.4.3 Histogramas

Clase 9, Mario Sánchez Domínguez

Examen 2

5.0 Programación

5.1 Estructuras de control

5.1.1 Condicionales (if/else)

5.1.2 Bucles (while/for)

Clase 10, Mario Sánchez Domínguez

5.2 Funciones personalizadas

5.3 Repaso programación

Clase 11, Mario Sánchez Domínguez

6.0 Modelos

6.1 Distribuciones estadísticas

6.2 Modelos lineales sencillos

Clase 12, Mario Sánchez Domínguez

6.3 Modelos multifactoriales

6.4 Pruebas estadísticas básicas

Clase 13, Mario Sánchez Domínguez

7.0 Temas avanzados y repaso

7.1 Instalación de librerías

7.1.1 Librería usadas frecuentemente

7.2 Objetos avanzados

7.2.1 Imágenes

7.2.1 Rasters

7.2.3 Polígonos

7.3 Repaso del curso

Clase 14, Mario Sánchez Domínguez, Federico Castro Monzón

8.0 Desarrollo del proyecto personal

Clase 15 Mario Sánchez Domínguez, Federico Castro Monzón

8.1 Trabajo con asesoría

Clase 16

reposición

Bibliografía

R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>. Higgins, P., (2022) Reproducible medical research with R. URL https://www.bookdown.org/pdr_higgins/rmrwr/#helpful-tools Golemund, G., & Wickham, H. (2017). R for data science. O'Reilly Media.