



4 créditos

2023-2023

área:
Fisicomatemáticas

Turno: indistinto

max. estudiantes: 20

9 semestre



SEMESTRAL

Temas Selectos de Matemáticas (Bioestadística mediante lenguaje R)

Profesor: Dr. Luis Escalera Vázquez

HORARIO

TEORÍA y
PRÁCTICA

martes y jueves de 9-11 am

LUGAR:

Edificio X

OBJETIVO

Identificar, relacionar e interpretar conceptos y métodos de minería de datos, selección de análisis estadísticos y aplicarlo a modelos y procesos biológicos mediante el uso de programación R.

Información adicional:

Se realizarán procesos computacionales y uso de plataformas virtuales, por lo que se recomienda que el alumno esté familiarizado con el uso de hojas de cálculo y búsqueda de información en internet.

Curriculum brevis

Nombre: Luis Humberto Escalera Vázquez

RFC: EAVL800802Q48

Correo electrónico: lhescalera@gmail.com / humberto.vazquez@umich.mx

Google Scholar: <https://scholar.google.com/citations?user=uQ30KK0AAAAJ&hl=en>

He incursionado en el campo de la Ecología Acuática utilizando principalmente a los peces y anfibios como modelo de estudio. He incorporado diferentes disciplinas tales como la Ecología de Comunidades, Acuicultura, Manejo y Conservación de Recursos Naturales y Ecosistemas Acuáticos. Mi formación académica es: Licenciatura en Biología, Universidad Michoacana de San Nicolás de Hidalgo (1998-2003); Maestría (2004-2006) y Doctorado (2007-2011), Instituto de Biología, Universidad Nacional Autónoma de México, estos últimos obtenidos con Mención Honorífica. Realicé una estancia postdoctoral en el extranjero, en el Instituto Nacional para Estudios Ambientales, Japón (2012-2013) y una como Investigador Asociado en la Universidad Autónoma de Campeche (2014-2015). Actualmente funjo como Profesor-Investigador Titular “A” en la Facultad de Biología-UMSNH y formo parte del Laboratorio de Biología Acuática. Cuento con más de 28 publicaciones internacionales y soy miembro del Sistema Nacional de Investigadores (SNI) como nivel 1. He generado investigación relevante sobre los cambios en las comunidades acuáticas relacionadas a la modificación de las condiciones ambientales ya sea de manera natural o por actividades antrópicas. Específicamente, he estudiado procesos ecológicos en peces y anfibios como el uso de recursos alimenticios por especies endémicas, desplazamiento competitivo por especies exóticas, factores ambientales que estructuran las comunidades en humedales, y distribución potencial y efecto del cambio climático en anfibios. También he participado en proyectos multidisciplinarios e interinstitucionales, en donde se integraron diferentes problemáticas ambientales, con el objetivo de generar y transferir el conocimiento y la tecnología a diferentes sectores.



**UNIVERSIDAD MICHOACANA DE
SAN NICOLÁS DE HIDALGO**

FACULTAD DE BIOLOGÍA



CURSO OPTATIVO

Programa para la materia: Temas Selectos de Matemáticas (Bioestadística mediante lenguaje R)

Fecha de Inicio: febrero 2023 (semestre 23-23)

Carga horaria: 4 horas semanales

Horas totales al semestre: 64

Créditos: 4

Área académica: Física y Matemáticas

Lugar: Laboratorio de cómputo (Edificio X-Ciudad Universitaria)

Modalidad: Presencial

Fecha de elaboración: 30 de septiembre del 2022

Fecha de última revisión y actualización: 30 de septiembre del 2022

Participantes en la Elaboración: Dr. Luis H. Escalera Vázquez

Participantes en el Desarrollo: Dr. Luis H. Escalera Vázquez

Información al correo electrónico: humberto.vazquez@umich.mx

Participantes en la Revisión: Dr. Rodolfo Pérez Rodríguez

Dr. Javier Ponce Saavedra

M. en C. Jennifer López Chacón

Perfil profesional del profesor: Especialista en el análisis de comunidades biológicas y en el manejo de datos ecológicos. Conocimiento y uso avanzado de análisis estadísticos univariados y multivariados; manejo de información biológica y su relación con diferentes procesos ecológicos.

Programas y sistemas operativos computacionales requeridos

Windows, Mac, Linux (Ubuntu) Microsoft EXCEL

R versión 4.1.2 o posterior: disponible en: <http://cran.r-project.org/>

Interface Gráfica, se sugiere RStudio: disponible en <http://www.rstudio.com/ide/download/>

R en la nube: disponible en <https://rstudio.cloud/>

Introducción

Dentro de los retos que enfrentan los profesionales en Ciencias Biológicas (ej. Biología y ecología), se encuentra principalmente, el dar respuestas concretas a problemáticas ambientales a través del análisis de la naturaleza, mediante la descripción e interpretación de las relaciones entre las especies (o conjuntos de especies) con el ambiente. Por lo que contar con las herramientas analíticas que permitan el uso adecuado de la información adquirida en campo, y a partir de la interpretación precisa de los datos, tomar decisiones sobre los resultados de la investigación. Dentro de las Ciencias Ambientales, el estudio de las comunidades se ha convertido en una pieza importante, dada su inherente relación con las necesidades del siglo XXI, y con la importancia que ésta tiene para el sustento humano y para el mantenimiento de la biodiversidad. De los ecosistemas, se puede adquirir una gran variedad de datos que nos proporcionen respuestas y relaciones de las comunidades biológicas con el ambiente.

Uno de los principales retos en las Ciencias Ambientales es dar a conocer la información de una manera adecuada y precisa. Gracias a la facilitación en el uso de computadoras (ej. costos, formas e información de uso de diferentes programas, etc.) en las últimas dos décadas, la exploración (minería), manejo y procesamiento de datos ha facilitado el análisis de los mismos, así como nos permite realizar análisis más complejos utilizando grandes bases de datos. Esto sugiere, que el utilizar y conocer herramientas estadísticas y programas computacionales actualizados es esencial para las demandas actuales en el análisis de datos biológicos, para así dar respuestas más precisas sobre como la naturaleza trabaja.

Por ello, el conocer las bases de los análisis estadísticos y matemáticos desde una perspectiva biológica, permite interpretar de manera más precisa los patrones y respuestas a fenómenos naturales, y generar propuestas de manejo, conservación e investigación de los diferentes componentes de la naturaleza.

El presente curso está diseñado para utilizar como herramienta básica la Bioestadística, aplicando modelos matemáticos a los procesos biológicos y ecológicos, ampliando el conocimiento sobre: diseños y procedimientos experimentales, clasificación de información, inferencias y síntesis de resultados.

En el presente curso el alumno desarrollará procedimientos de minería de datos, para decidir los análisis estadísticos a utilizar con un enfoque multivariado, asumiendo que conoce y utiliza herramientas de Bioestadística básica. A lo largo del curso se desarrollarán ejercicios de práctica con un sistema de ayuda, glosario de términos y bibliografía de referencia.

Dado que el programa R es una herramienta actual (y actualizable), accesible y práctica, la cual cuenta con foros internacionales de discusión, y que su versatilidad se ha extendido a cualquier área que requiera de análisis matemáticos y del manejo de bases de datos, se toma como base el conocer los principios del funcionamiento de dicho programa, así como se introducirá e impulsará al alumno para que genere sus propios análisis, bases de datos y sistemas de programación.

Objetivo

Identificar, relacionar e interpretar conceptos y métodos de minería de datos, selección de análisis estadísticos y aplicarlo a modelos y procesos biológicos mediante el uso de programación R.

Contenido programático:

I. DISEÑO EXPERIMENTAL Y DE MUESTREO (4 horas)

Objetivo: Identificar el diseño de muestreo o experimental, así como las variables que se ajusten a una pregunta de investigación

- 1.1 Identificación de la pregunta de investigación
- 1.2 Evaluación logística y técnicas de medición
- 1.3 Decisión de variables y factores en experimento y campo
- 1.4 Identificación de interacciones, agrupamientos y número de variables y factores
- 1.5 Número mínimo de muestra y réplicas
- 1.6 Identificación de diseños y análisis completos e incompletos (en bloques)
- 1.7 Identificación de error en la toma de muestras y error estadístico
- 1.8 Generación de resultados potenciales y búsqueda de análisis estadísticos

II. INTRODUCCIÓN AL LENGUAJE R (6 horas)

Objetivo: Conocer el lenguaje fundamental de todos los programas computacionales y traducirlo al lenguaje biológico tomando como modelo de estudio el lenguaje R

- 2.1 Conocimientos básicos sobre R Instalación de R
- 2.2 Paquetes (básicos y adicionales) Inicio sesión en R
- 2.3 Naturaleza y tipo de datos Funciones generales

III. DATOS UNIVARIADOS (4 hora)

Objetivo: Identificar la naturaleza de datos y distribución desde una perspectiva biológica.

- 3.1 Identificación de datos cualitativos
- 3.2 Identificación de datos cuantitativos
- 3.3 Identificación de distribuciones estadísticas (poisson, normal, binomial)

IV. DIMENSIONES EN BASES DE DATOS (4 horas)

Objetivos: Identificar dentro del lenguaje R las diferentes estructuras de bases de datos

- 4.1 Vectores Definición, tipos y usos
- 4.2 *Arrays* y matrices Definición, tipos y usos

V. GRÁFICOS (6 horas)

Objetivo: Conocer los diferentes gráficos que se pueden generar en el lenguaje R y su interpretación

- 5.1 Parámetros gráficos de la función “plot”
- 5.2 Elementos complementarios de un gráfico
- 5.3 Interpretación y uso de gráficos en minería de datos
- 5.4 Formatos para guardar los gráficos

VI. DATOS MULTIVARIADOS (4 horas)

Objetivo: Identificar desde una perspectiva biológica efectos aditivos y sinérgicos al considerar bases de datos con más de una variable y su procesamiento en lenguaje R

- 6.1 Estructura de datos (*data frames*)
- 6.2 Identificación y acceso a los datos
- 6.3 Uso de: *apply*, *tapply*, *lapply*
- 6.4 Manejo y conversión de listas

VII. INTEGRACIÓN DE ARCHIVOS PARA ANÁLISIS DE DATOS (6 horas)

Objetivo: Reconocer los diferentes tipos de archivos de bases de datos y su lectura en el lenguaje R

- 7.1 Lectura y manipulación de datos
- 7.2 Importar y exportar datos
- 7.3 Guardar y leer datos

VIII. CREAR FUNCIONES (6 horas)

Objetivo: Crear funciones estadísticas en lenguaje R con fundamento biológico y que sean aplicables a sistemas de estudio

- 8.1 ¿Qué es una función?
- 8.2 Consideraciones y argumentos

IX. MODELOS LINEALES (6 horas)

Objetivo: Identificar e interpretar diferentes modelos lineales aplicables a sistemas biológicos y generar análisis en lenguaje R

- 9.1 Regresión lineal
- 9.2 Regresión múltiple
- 9.3 Análisis de la varianza (ANOVA)
- 9.4 Modelos lineales Generalizados (GLMs)

X. ANÁLISIS MULTIVARIADO (6 horas)

Objetivo: Identificar diferentes análisis multivariados, así como reconocer la diferencia biológica entre variable y factor.

- 10.1 Análisis de varianza multivariado (MANOVA)
- 10.2 Análisis de componentes principales (PCA)
- 10.3 Análisis de correspondencia tendencial (DCA)
- 10.4 Análisis Discriminante (AD)
- 10.5 Análisis no- métricos multidimensionales (NMDS)
- 10.6 Análisis de similitud (ANOSIM)
- 10.6 Análisis indirectos o integrativos

XI. ALGUNOS ANÁLISIS CLÁSICOS Y BÁSICOS DE COMUNIDADES (4 horas)

Objetivo: Dar una perspectiva general de análisis estadísticos diseñados especialmente a partir de preguntas biológicas y realizarlos con librerías en el programa R

- 11.1 Diversidad
- 11.2 Rarefacción
- 11.3 Permutaciones: uso en ecología y parte integral en análisis estadísticos

XII. INTEGRACIÓN DE ANÁLISIS E INTERPRETACIÓN DE DATOS PARA EL ESTUDIO DE COMUNIDADES (4 horas)

Objetivo: Identificar el uso de los resultados de diferentes análisis y el uso e integración con análisis complementarios desde una perspectiva biológica y matemática mediante el lenguaje R.

- 11.1 ¿Qué nos dicen los resultados de todos los análisis?

XIII. UN PASO MÁS EN LA EXPLORACIÓN Y MANEJO DE R (4 horas)

Objetivo: Mostrar las diferentes herramientas de actualización y generación de nuevos análisis biológicos y cómo implementarlos en el lenguaje R, ya sea mediante librerías, funciones y comandos.

- 12.1 Búsqueda de análisis, librerías y reconocimiento de datos
- 12.2 Futuro y perspectivas de la programación en ciencias biológicas usando R

PRÁCTICAS DE LABORATORIO Y CAMPO

El curso se desarrollará en su totalidad en el aula de cómputo del edificio X de la Facultad de Biología, Universidad Michoacana de San Nicolás de Hidalgo. Se desarrollarán diferentes ejercicios estadísticos con un enfoque 100% biológico.

METODOLOGÍA Y DESARROLLO GENERAL DEL CURSO.

A lo largo del curso el profesor explicará los diferentes módulos mediante ejemplos biológicos a través de exposiciones frente al grupo. Se compartirá información útil para el desarrollo del curso como son: bases de datos, librerías estadísticas, páginas web de soporte, entre otras.

Los alumnos desarrollarán ejercicios de cómputo en los cuales se plantearán preguntas dentro del marco de diferentes disciplinas biológicas. Dichos ejercicios se resolverán en clase o como tareas.

SISTEMA GENERAL DE EVALUACIÓN.

EVALUACIÓN DE LA PARTE TEÓRICA

Participaciones diarias	10%
Tareas semanales	10%
Numero de exámenes parciales:	
1° (unidades I a la VI)	20%
2° (Unidades VII a la XII)	20%
SUMA TOTAL	60 %

EVALUACIÓN DE LA PARTE PRÁCTICA.

En las Prácticas de Laboratorio de Cómputo se considerarán las asistencias, entrega de reportes y presentación de material preparado (reportes, explicación de resultados y gráficas, presentaciones por parte de los alumnos, etc).

Generación de resultados con base en las prácticas por clase	10%
Asistencia y entrega de reportes	10%
Presentación final de un proyecto de investigación	20 %
SUMA	40%

Al final del curso solo se obtendrá una calificación, para que el promedio pueda realizarse es necesario que ambas partes (teoría y práctica) sean aprobatorias.

$$\text{CALIFICACIÓN FINAL} = \text{TEORÍA} + \text{PRÁCTICA} / 2$$

SALIDA A CAMPO

No aplica

CORRELACIÓN CON OTRAS MATERIAS.

Estadística, Ecología, Bioinformática, Programación, diseño de muestreo y experimental, Matemáticas, Álgebra, trigonometría.

CRONOGRAMA DE ACTIVIDADES

Actividades en aula (laboratorio de cómputo; Teoría y práctica)

Unidad 1. semana 1.

Unidad 2. semana 2 y 3.

Unidad 3. semana 3 y 4.

Unidad 4. semana 4 y 5.

Unidad 5. semana 5 a la 7.

Unidad 6. semana 8.

Unidad 7. semana 8 y 9.

Unidad 8. semana 9 y 10.

Unidad 9. semana 10 y 11.

Unidad 10. semana 11 a la 13.

Unidad 11. semana 14 y 15.

Unidad 12. semana 16.

BIBLIOGRAFÍA

Libros:

Buttigieg P.L., Ramette A (2014) A Guide to Statistical Analysis in Microbial Ecology: a community-focused, living review of multivariate data analyses. *FEMS Microbiol Ecol.* 90: 543–

550. Stevens, M. H. (2010) *A Primer of Ecology with R*. Springer

Everitt, B. S. & I. Hothorn (2010) *A Handbook of Statistical Analyses Using R*, 2° edn. CRC Press.

Venables, W. N. & B. D. Ripley (2002) *Modern Applied Statistics with S*. Springer

Paradis, E. (2002) *R para Principiantes*. Institut des Sciences de l' Evolution. Universit Montpellier II, France.

Horton, N. J. & K. Kleinman (2011) *Using R for Data Management, Statistical analysis, and Graphics*. CRS Press, Boca Raton, FL.

Bolker, B. (2007) *Ecological Models and Data in R*. Princeton University Press. Morin, P.J.

(2005) *Community Ecology*, Blackwell Science, United States of America.

Tilman, D. (1982) *Resource Competition and Community Structure*, Princeton University Press, Princeton.

Magurran, A.E. (2003) *Measuring Biological Diversity*, Wiley-Blackwell. Ricklefs, R.E. & Miller,

G. (1999) *Ecology*, W. H. Freeman.

Begon, M., Townsend, C.R. & Harper, J.L. (2006) *Ecology: From Individuals to Ecosystems*, Wiley- Blackwell.

Artículos:

Chase, J.M. (2003) Community assembly: when should history matter? *Oecologia*, 136, 489–498.

Colwell, R. K., A. Chao, N. J. Gotelli, S.-Y. Lin, C. X. Mao, R. L. Chazdon & J. T. Longino

(2012).

Models and estimators linking individual-based and sample-based rarefaction, extrapolation and comparison of assemblages. *Plant Ecology* 5:3-21.

Gotelli, N.J. & Colwell, R.K. (2001) Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters*, 4, 379 - 391.

Baber, M.J., Childers, D.L., Babbitt, K.J. & Anderson, D.H. (2002) Controls on fish distribution and abundance in temporary wetlands. *Canadian Journal of Fisheries and Aquatic Sciences*, 59, 1441– 1450.

Escalera-Vázquez Luis H., Nancy Calderón-Cortés and Luis Zambrano-González (2017). Fish population responses to hydrological variation in a seasonal wetland in southeast México. *Neotropical Ichthyology*

Escalera-Vázquez L.H., Omar Domínguez-Domínguez, Demián Hinojosa-Garro & Luis Zambrano (2016). Changes in diet, growth and survivorship of the native Tequila Splitfin *Zoogoneticus tequila* in co-occurrence with the non-native Shortfin Molly *Poecilia mexicana*. *Fundamental and Applied Limnology*.

Escalera-Vázquez, L.H. & Zambrano, L. (2010) The effect of seasonal variation in abiotic factors on fish community structure in temporary and permanent pools in a tropical wetland. *Freshwater Biology*, 2557-2569.

Kobza, R.M., Trexler, J.C., Loftus, W.F. & Perry, S.A. (2004) Community structure of fishes inhabiting aquatic refuges in a threatened Karst wetland and its implications for ecosystem management. *Biological Conservation*, 116, 153-165.

Pazin, V., F. V., Magnusson, W., Zuanon, J. & Mendoca, F., P. (2006) Fish assemblages in temporary ponds adjacent to ‘terra-firme’ streams in Central Amazonia. *Freshwater Biology*, 51, 1025–1037.

Súarez, Y.R., Júnior, M.P. & Catella, A.C. (2004) Factors determining the structure of fish communities in Pantanal lagoons. *Fisheries Management and Ecology*, 8, 173–186.

Tokeshi, M. (1993) Species abundance patterns and community structure. *Advances in Ecological Research*, 24, 111-186.

Magoulick, D.D. & Kobza, R.M. (2003) The role of refugia for fishes during drought: a review and synthesis. *Freshwater Biology*, 48, 1186–1198.

Palma, A. (2010) ¿Nicho, teoría neutral, o una alternativa emergente? *Ecología Austral*, 20, 63-69.

Kushlan, J.A. (1976) Environmental Stability and Fish Community Diversity. *Ecology*, 57, 821-

825. Southwood, T.R.E. (1977) Habitat, the templet for ecological strategies? *Journal of Animal Ecology*,