

## INTRODUCCION A LOS MODELOS MULTIDIMENSIONALES EN R

### OBJETIVO GENERAL

El alumno conocerá los fundamentos teóricos y aplicados de los modelos multidimensionales en estadística (también conocida como estadística multivariada) para el análisis de datos, poniendo especial énfasis en los aspectos interpretativos y de estrategia de análisis (más que en los estrictamente operativos). Se favorecerá la noción del uso eficiente de las herramientas cuantitativas en el campo de las distintas disciplinas que estudian los sistemas acuáticos, reforzando los aspectos metodológicos de la investigación científica y adquiriendo habilidades en el uso de plataformas computacionales de actualidad, particularmente en el lenguaje de programación R (R-Project).

### OBJETIVOS ESPECÍFICOS

- Que el alumno distinga las situaciones en las que es indispensable utilizar modelos estadísticos de respuesta múltiple como la mejor estrategia para la resolución de problemas en el área de las ciencias acuáticas.
- Que el alumno comprenda la problemática asociada a los juegos de datos multidimensionales, e identifique distintas estrategias numéricas y de representación gráfica que permitan la descripción adecuada de los fenómenos estudiados en las ciencias acuáticas.
- Que el alumno conozca los distintos métodos de clasificación y ordenación usados para explorar juegos de datos multidimensionales, e identifique ventajas y desventajas en el contexto de distintos problemas en el ámbito de las ciencias acuáticas.
- Que el alumno comprenda y aplique los métodos de inferencia estadística para poner a prueba hipótesis sobre datos multivariados, e identifique los alcances y limitaciones de estos procedimientos en el contexto de las ciencias acuáticas.
- Que el alumno comprenda y aplique distintas técnicas de re-muestreo para la solución de problemas implícitos en los métodos confirmatorios aplicados a juegos de datos multidimensionales en el ámbito de las ciencias acuáticas.
- Que el alumno conozca y aplique el lenguaje de programación en R para los cálculos asociados a las técnicas exploratorias y confirmatorias revisadas en las clases teóricas utilizando juegos de datos realistas, ya sea provenientes de proyectos de investigación, o bien, simulados para resaltar rasgos comunes y problemáticas frecuentes en el ámbito de las ciencias acuáticas.
- Que el alumno compare las rutinas de programación en R con el uso y las salidas de algunos paquetes de cómputo estadístico, frecuentemente usados en estudios de las ciencias acuáticas y otras áreas afines.

### DIDACTICA / DINAMICA EN CLASE

Reconociendo la importancia de que los estudiantes de nivel superior comprendan las ideas básicas relativas al diseño de experimentos y la colección, manejo, curación y análisis cuantitativo de datos, existe una amplia oferta de cursos en probabilidad y estadística. Sin embargo, el abordaje de los cursos tradicionales se enfoca en el desarrollo de habilidades técnicas

(algorítmicas) versus la comprensión de los métodos e interpretación de resultados. Los paquetes tecnológicos comúnmente usados (o bien, la forma como son usados) muchas veces obstruyen, más que facilitan el aprendizaje de conceptos estadísticos y sus aplicaciones. En general, entre los estudiantes se produce frustración, desinterés y una respuesta de rechazo acompañada de la convicción de que no son capaces de comprender los conceptos relativamente complejos de probabilidad, estadística y diseño experimental.

El presente curso utiliza métodos pedagógicos que enfatizan el desarrollo del razonamiento estadístico sobre la aplicación automatizada de pruebas de hipótesis. Para contribuir al cumplimiento de este objetivo, se consideraron los siguientes elementos:

- Asegurar una adecuada contextualización del problema de investigación mediante el uso de ejemplos concretos con datos y situaciones de investigación reales. Esto facilita la traducción del enunciado del problema a la formulación matemática (abstracta) del modelo estadístico, así como la interpretación lógica de los resultados estadísticos en el contexto del problema.
- El uso de la tecnología computacional a través de actividades de programación abiertas basadas en teorías constructivistas del conocimiento, y más específicamente en el paradigma del "construccionismo". Con estas herramientas didácticas, los estudiantes pueden explorar y construir ideas y conceptos a través de rutinas programáticas elaboradas por ellos mismos, ya sea *di novo* o a través de modificaciones de otras ya construidas.
- El uso del lenguaje gráfico como instrumento de transnumeración. El paso de un registro de representación a otro facilita la revelación de la información novedosa en cada paso y construye una forma básica de razonamiento estadístico.
- La diversificación de métodos de evaluación y autoevaluación que aseguren una retroalimentación oportuna, y promuevan un aprendizaje significativo a través de la prueba y el error.

En años recientes se ha generalizado el uso de R (de distribución gratuita) para el análisis cuantitativo de datos de diversa índole. Al ser un programa y un lenguaje de programación a la vez, R contempla la manipulación de objetos y su representación, e implica representaciones ejecutables a través de una serie de comandos intuitivos, cuya sintaxis es sencilla, repetible y consistente. Mediante su uso como plataforma de análisis en este curso, se proporciona al estudiante las herramientas programáticas para llevar a cabo los procedimientos estadísticos aprendidos, enseñándole a interpretar correctamente las salidas tanto numéricas como gráficas de las distintas rutinas. El uso combinado de plataformas de programación para el cómputo de los modelos estadísticos y la intervención continua con fines interpretativos a lo largo del procedimiento analítico, diversifica las competencias de los alumnos, y potencializa la aplicación de ésta importante herramienta analítica.

Con esto en mente, se elaboraron actividades de programación computacional en R, que van acompañadas de una serie de instrucciones, guías, ejemplos y tareas de programación. Al final de cada sesión ponemos a disposición de los estudiantes diversas preguntas para la reflexión y soluciones comentadas. La comprensión de los modelos estadísticos se facilita a través de la creación de objetos (virtuales) que los representan, por lo que los estudiantes están invitados a

familiarizarse con el lenguaje de programación y sus librerías de funciones. Además de fortalecer la enseñanza del razonamiento estadístico, esto representa una plusvalía, especialmente ante el uso cada vez más generalizado de esta plataforma de acceso libre para usuarios en todas las áreas de conocimiento.

TEMARIO	No. de horas	
	Teor	Prac
UNIDADES TEMÁTICAS		
Unidad I Conceptos preliminares <ol style="list-style-type: none"> <li>1. Revisión de estadística univariada.</li> <li>2. La variable de respuesta única y múltiple.</li> <li>3. Razones para usar modelos multidimensionales (cuándo usarlos).</li> <li>4. Notación estadística.</li> <li>5. Revisión de álgebra matricial.</li> </ol>	2	6
Unidad II Descripción de patrones con datos multi-D <ol style="list-style-type: none"> <li>1. El problema de las n-dimensiones.</li> <li>2. Medidas de asociación: R análisis. Matriz VCV y de correlación. Distancia <math>X^2</math>.</li> <li>3. Medidas de asociación: Q análisis. La distancia/similitud multidimensional.</li> <li>4. Medidas de distancia para datos binarios, discretos ("multi-state") y continuos.</li> <li>5. Métricas y medidas. El espacio euclidiano.</li> </ol>	2	8
Unidad III Métodos exploratorios: Clasificación <ol style="list-style-type: none"> <li>1. Clusters</li> <li>2. Tipos de dendrogramas y métodos de agrupamiento.</li> <li>3. Métodos paramétricos y no paramétricos (k-clustering).</li> <li>4. Árboles de regresión/clasificación</li> <li>5. Resumen. Ventajas y desventajas (limitaciones) de la clasificación</li> </ol>	2	6
Unidad IV Métodos exploratorios: Ordenación (1er encuentro). <ol style="list-style-type: none"> <li>1. Ordenación como una rotación rígida de los ejes.</li> <li>2. Análisis de Componentes Principales (PCA).</li> <li>3. Obtención de la matriz de VCV. Descomposición de eigenvalores.</li> <li>4. Interpretación numérica y gráfica de un PCA.</li> <li>5. Visualización e interpretación gráfica de un PCA.</li> </ol>	2	8

<p>Unidad V Métodos exploratorios: Ordenación (2o encuentro)</p> <ol style="list-style-type: none"> <li>1. Escalamiento multidimensional (PCoA en R: funciones definidas).</li> <li>2. Interpretación numérica y gráfica.</li> <li>3. Evaluación de la configuración (“Shepard plots” y estrés).</li> <li>4. Escalamiento multidimensional no-métrico (NMDS en R: funciones definidas).</li> <li>5. Interpretación. Evaluación de la configuración (“Shepard plots” y estrés).</li> <li>6. Resumen. Ventajas y desventajas (limitaciones) de la ordenación.</li> </ol>	2	8
<p>Unidad VI: Métodos de inferencia estadística: Modelos Lineales Multi-D</p> <ol style="list-style-type: none"> <li>1. Revisión de GLM univariados. Estimativa de parámetros.</li> <li>2. GLM con matrices. Resolviendo para Beta.</li> <li>3. GLM multidimensional: MANOVA. El estadístico F. Pruebas post-hoc (<math>T^2</math>).</li> <li>4. GLM multidimensional: Regresión. Estadísticos. Visualización.</li> <li>5. Requisitos de los modelos GLM y consecuencias de violarlos.</li> </ol>	2	8
<p>Unidad VII: Métodos de inferencia estadística: Técnicas de re-muestreo</p> <ol style="list-style-type: none"> <li>1. La limitante de las distribuciones teóricas.</li> <li>2. El raciocinio de las técnicas de re-muestreo. La prueba exacta de Fisher.</li> <li>3. Tipos de re-muestreo. Métodos de Bootstrap, Jackknife y Monte Carlo</li> <li>4. Permutación irrestricta y restringida (de datos crudos) y de residuales.</li> <li>5. Abordaje basado en la distancia multidimensional. Soluciones al problema <i>Behrens-Fisher</i>.</li> </ol>	2	8
	14	52
		66