

## MODELOS LINEALES GENERALIZADOS EN R

### OBJETIVO GENERAL

El alumno conocerá los fundamentos teóricos y aplicados de los Modelos Lineales Generalizados para el análisis de datos, poniendo especial énfasis en los aspectos interpretativos y de toma de decisiones (más que en los estrictamente operativos). Se favorecerá la noción de uso eficiente de las herramientas cuantitativas en el campo de las distintas disciplinas que estudian los sistemas acuáticos, reforzando los aspectos metodológicos y adquiriendo habilidades en el uso de plataformas computacionales de actualidad.

### OBJETIVOS ESPECÍFICOS

Al finalizar el curso el alumno:

- Conocerá y aplicará los conceptos fundamentales de la regresión lineal simple y múltiple, incluyendo tanto el enfoque del análisis exploratorio como el de diseño de experimentos de medición y manipulación, ambos en el contexto de los sistemas acuáticos.
- Revisará el raciocinio implícito en las pruebas de hipótesis, incluyendo reflexiones sobre la probabilidad de errores tipo I y II.
- Conocerá y aplicará el lenguaje de programación en R para los cálculos asociados a nociones básicas de los modelos lineales y sus extensiones hacia juegos de datos con distribuciones subyacentes distintas a la distribución normal.
- Comparará las rutinas de programación en R con los resultados de algunos paquetes de cómputo estadístico.

### 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. El problema de la variabilidad y la inferencia estadística <ol style="list-style-type: none"> <li>Tres marcos conceptuales para las herramientas estadísticas</li> <li>Poblaciones y muestras. La aleatoriedad y la independencia.</li> <li>Concepto de probabilidad. Distribuciones de variables aleatorias.</li> <li>Construcción de una prueba estadística.</li> <li>Componentes de una prueba. Tipos de error I y II.</li> </ol>	2	6
UNIDAD II. Modelos generales lineales (LM) <ol style="list-style-type: none"> <li>Regresión lineal bi-variada. La función <math>\chi</math> indica causalidad?</li> <li>El modelo aditivo. Coeficientes del modelo y su interpretación.</li> <li>Componentes de la varianza. Análisis de varianza y regresión</li> <li>Significancia vs indicadores de ajuste. Coeficiente de determinación</li> </ol>	2	6
UNIDAD III. Análisis de Residuales y Exploratorio de Datos <ol style="list-style-type: none"> <li>Supuestos de GLM y su violación.</li> <li>Validación gráfica y analítica del modelo (AR).</li> <li>Alternativas a priori. Diseño y/o ejecución de experimentos.</li> <li>Alternativas a posteriori. Re-expresión de los datos.</li> <li>Análisis exploratorio de datos (AED).</li> </ol>	2	8
UNIDAD IV. Regresión múltiple y ANCOVAR <ol style="list-style-type: none"> <li>Los coeficientes parciales de la regresión.</li> <li>Significancia de los parámetros en RM.</li> <li>Interacciones y su interpretación.</li> <li>Selección del modelo ideal (MST): criterios y procedimientos.</li> <li>AR y AED en la regresión múltiple.</li> <li>ANCOVAR</li> </ol>	2	8
UNIDAD V. Extensiones de los modelos lineales <ol style="list-style-type: none"> <li>La estructura aleatoria de los modelos.</li> <li>Distribuciones no-normales: modelos lineales generalizados (GLM)</li> <li>Regresión de Poisson</li> <li>El parámetro <math>\phi</math> (sobre y sub-dispersión en GLM)</li> <li>Regresión Logística</li> <li>Heterogeneidad: estructuras de varianza (Introducción a GLMM)</li> </ol>	2	8
	10	36
<b>TOTAL</b>		<b>46</b>