

DISEÑO EXPERIMENTAL Y ANÁLISIS DE DATOS EN R

OBJETIVO GENERAL

El alumno conocerá los fundamentos teóricos y aplicados del diseño experimental y 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 adquiera destreza en la aplicación de conceptos relativos a la variabilidad de los sistemas naturales, su cuantificación y su control a través del diseño experimental.

- Que el alumno desarrolle la capacidad para discernir entre distintas alternativas de análisis cuantitativo, tomando en cuenta tanto la naturaleza estadística de los datos como los objetivos específicos de una investigación.
- Que el alumno adquiera habilidades de interpretación gráfica, y transmisión escrita y oral de las ideas que le permitan poner de manifiesto su capacidad argumentativa, especialmente para reconocer los alcances y limitaciones de sus resultados, e identificar los caminos por donde puede continuar una investigación.
- Que el alumno comprenda y aplique los métodos de inferencia estadística para poner a prueba hipótesis sobre fenómenos complejos en el contexto 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 básicos 1. Definición y utilidad. 2. Tres marcos estadísticos conceptuales 3. El problema de la variabilidad. 4. Poblaciones y muestras. Parámetros y variables. 5. Tipos de variables y escalas de medición. 6. Distribuciones de frecuencia.	2	6
UNIDAD II. Probabilidad 1. Conceptos básicos de probabilidad 2. Distribuciones de probabilidad: Binomial 3. Distribuciones de probabilidad: Normal 4. Teorema Central del Límite	2	6
UNIDAD III. Diseño experimental y pruebas de hipótesis (simples) 1. Construcción de una prueba estadística. 2. Errores tipo I y II. 3. Procedimientos básicos para un correcto diseño experimental 4. Pruebas de hipótesis (simples). 5. Distribución de t-Student.	2	6
UNIDAD IV. Análisis de varianza. 1. Concepto general. Racional detrás del ANOVA 2. Diseños con factores fijos y aleatorios. 3. Diseño completamente aleatorizado. 4. El modelo aditivo y la tabla de análisis de varianza. 5. Comparaciones de medias: contrastes vs comparaciones post hoc.	2	6
UNIDAD V. Diseños de ANOVA anidados. 1. Racional detrás del uso de bloques. 2. Diseño completamente aleatorizado con bloques al azar. 3. El modelo aditivo y la tabla de análisis de varianza. 4. Comparaciones múltiples. 5. Diseños anidados complejos.	2	6
UNIDAD VI. Diseños factoriales. 1. Diseño balanceado con efectos fijos. 2. El modelo aditivo y la tabla de análisis de varianza. 3. Interpretación de interacciones. 4. Diseño balanceado con efectos aleatorios y efectos mixtos. 5. La tabla de análisis de varianza. 6. Algunos diseños experimentales frecuentemente usados.	2	6
UNIDAD VII. Modelos lineales (LM). 1. Medidas de asociación. Correlación lineal. 2. Coeficiente de Pearson y pruebas de significancia. 3. Regresión lineal bi-dimensional. Coeficientes y el modelo aditivo. 4. Componentes de la varianza. Análisis de varianza y regresión 5. Significancia vs indicadores de ajuste. Coeficiente de determinación.	2	6

<p>UNIDAD VIII. Validación de los modelos lineales.</p> <ol style="list-style-type: none"> Supuestos de los modelos. Verificación gráfica y analítica de los supuestos. Alternativas de diseño y/o ejecución de los experimentos (a priori). Alternativas de análisis de los resultados (a posteriori). La estructura aleatoria de los modelos. Introducción a los Modelos Mixtos. 	2	6
<p>UNIDAD IX. Análisis exploratorio de datos</p> <ol style="list-style-type: none"> Detección de datos extremos: Análisis gráfico de la normalidad. Análisis gráfico de la homogeneidad de varianza. Detección de datos faltantes (librería "lattice"). Análisis gráfico de la relación funcional. Resumen y criterios para la toma de decisiones. 	2	6
<p>UNIDAD X. Introducción a la Regresión Múltiple</p> <ol style="list-style-type: none"> Análisis de Covarianza Coefficiente de determinación re-visitado AED (colinealidad) y validación de modelos de Regresión Múltiple. Herramientas para la selección del modelo. 	2	6
	20	60
TOTAL		80