



XS-0301: ESTADÍSTICA para BIÓLOGOS II, I Semestre 2021
Escuela de Estadística, Facultad de Ciencias Económicas, Universidad de Costa Rica

Docente: Hazel Quesada Leitón

Correo electrónico: hazel.quesadaleiton@ucr.ac.cr

Oficina: Escuela de Estadística, Facultad de Ciencias Económicas

Teléfono: 8611-4208

Horario de clases: L: 9:00-11:50 (teoría)

M: 14:00-15:50 (lab. Grupo 2)

J: 10:00-11:50 (lab. Grupo 1)

Horas de consulta: M: 10:00-11:40 Grupo 2

J: 13:00-14:40 Grupo 1

1. Descripción

Este curso se ubica en el bloque de asignaturas correspondientes al quinto semestre del plan de estudios de la carrera de Biología. El curso se orienta a dotar a los estudiantes de herramientas estadísticas útiles en diseño de experimentos, regresión y análisis multivariado. Se pretende desarrollar el sentido crítico del estudiante, fomentando su capacidad para afrontar y resolver problemas biológicos a partir de un planteamiento estadístico. Se pretende introducir una concepción estadística para la solución de problemas biológicos reales, orientada a obtener una respuesta satisfactoria.

El curso tiene un componente teórico donde se exponen las bases conceptuales de los métodos estadísticos y un componente práctico donde se expone al estudiante a datos reales para el análisis en el computador. Para la aplicación de los principios y métodos a cubrir en Estadística para Biólogos II es necesario dominar los conceptos fundamentales de la estadística descriptiva e inferencial.

- Requisitos: XS-0300 Estadística para biólogos I.
- Correquisitos: Ninguno.
- Horas: Dos sesiones por semana - 5 horas semanales (3 teoría y 2 práctica).
- Créditos: 5.

2. Objetivo General

Proveer a los estudiantes de la carrera de Biología del conocimiento conceptual y los instrumentos estadísticos avanzados para la investigación científica en su campo de acción.

3. Objetivos Específicos

Al finalizar el curso el estudiante tendrá criterio y conocimiento para:

- Entender y aplicar la técnica del análisis de variancia y los procedimientos de comparaciones múltiples para diferentes tipos de experimentos.
- Seleccionar un conjunto de variables para formar un modelo de regresión apropiado mediante un análisis detallado de todos los predictores disponibles.
- Llevar a cabo los diagnósticos del modelo de regresión ajustado.
- Comprender el modelo de regresión logística para respuestas dicotómicas.
- Seleccionar y aplicar la técnica multivariada apropiada cuando se tienen conjuntos de datos complejos.
- Usar adecuadamente el lenguaje de programación estadístico R para llevar a cabo las diferentes pruebas y técnicas estudiadas en el curso.



4. Contenidos

Módulo I: Diseño de experimentos.	
1.1	Análisis de variancia de una vía (diseño irrestricto aleatorio): <ol style="list-style-type: none"> Fuentes de variación: entre tratamientos y dentro de tratamientos. Comparación de fuentes de variación (tabla del ANDEVA). Prueba de homogeneidad de variancias. Pruebas robustas. Comparaciones múltiples: contrastes de Duncan, Tukey, etc.
1.2	Potencia o poder de la prueba: <ol style="list-style-type: none"> La importancia de la magnitud detectada por la prueba. Determinación del tamaño de muestra para pruebas de hipótesis. La potencia de una prueba realizada.
1.3	Otros diseños: <ol style="list-style-type: none"> Arreglos factoriales. Diseños con interacciones. Bloques aleatorios / medidas repetidas. Diseños anidados.
1.4	Pruebas no-paramétricas: <ol style="list-style-type: none"> Prueba de Kruskal Wallis. ANDEVA no-paramétrico con dos factores. Análisis de variancia de Friedman. Comparaciones múltiples.
Módulo II: Regresión.	
2.1	Correlación lineal entre dos variables.
2.2	El modelo de regresión múltiple: <ol style="list-style-type: none"> Estimación de los coeficientes. Valores ajustados y residuales. Prueba de hipótesis sobre los coeficientes. Estimación de respuesta media y predicción de nuevas observaciones. Coefficiente de determinación.
2.3	Diagnósticos.
2.4	Transformaciones, regresión polinomial y otras soluciones a problemas con los supuestos.
2.5	Regresión para respuestas de conteos (Poisson)
2.6	Regresión logística
Módulo III: Análisis multivariado.	
3.1	Análisis de componentes principales (PCA).
3.2	Análisis discriminante.
3.3	Análisis de conglomerados (Cluster).

5. Metodología

- El curso se impartirá de forma virtual, por tanto, es un curso de alta virtualidad, con clases sincrónicas y asincrónicas, desde la plataforma de medicación virtual y Zoom.
- Este curso es virtual para su versión I Semestre 2021. Se utilizará la plataforma institucional Mediación Virtual para colocar los documentos, presentaciones y vídeos del curso. Las clases asincrónicas se realizarán por medio de tareas, exámenes y foros, y las



- actividades sincrónicas por medio de la aplicación Zoom. Para las clases sincrónicas se compartirá con antelación la fecha y el enlace a utilizar, que permita al estudiante preparar su espacio físico y dispositivos necesarios.
- Se impartirán lecciones magistrales por parte del docente donde se explicarán los conceptos y sus aplicaciones.
- Cada persona utilizará su computadora para desarrollar 12 lecciones prácticas durante el semestre, las cuales corresponden a varios laboratorios por cada tema del programa.
- Durante las sesiones de laboratorio se utilizará el lenguaje de programación estadístico R para realizar ejercicios de la materia vista en clase.
- El profesor enviará los laboratorios con anticipación de tal forma que los estudiantes los hayan revisado previamente. Durante la sesión de laboratorio se evacuarán dudas sobre el uso del software y se profundizará en el análisis.
- Se asignarán tareas y ejercicios para asegurar la “puesta en práctica” de los conceptos estudiados. Se asignará una tarea por cada dos o tres laboratorios de tal forma que en total serán 4 tareas.
- Los estudiantes realizarán un trabajo de análisis de datos reales utilizando lo aprendido en el curso. Para esto presentarán un anteproyecto donde expondrán los objetivos del estudio y un plan de análisis el cual será criticado por el profesor y devuelto para que se realicen las mejoras sugeridas. Al final de curso se espera que los estudiantes escriban un artículo donde se describa detalladamente el procedimiento de análisis de los datos seleccionados.
- Se usarán como textos: Biostatistical Analysis de Zar y Análisis multivariado de Catena et al.

6. Evaluación

Se realizarán tareas y tres exámenes parciales en su computadora. Además los estudiantes presentarán un trabajo de análisis de datos. En el siguiente cuadro se presenta el desglose porcentual de la nota:

Rubro	Ponderación
Primer examen parcial	25%
Segundo examen parcial	25%
Tercer examen parcial	20%
Tareas individuales	15%
Trabajo de análisis de datos	15%
Total	100%

Si un estudiante faltara a algún examen por **causa justificada**, debe solicitar por escrito la reposición del examen indicando las razones de la ausencia, acompañada de los documentos justificantes. La misma debe entregarse ante el profesor que imparte el curso **a más tardar en cinco días hábiles del reintegro a lecciones**. No se recibirán exámenes ni tareas que incumplan con el tiempo establecido para su entrega.



7. Cronograma

	Módulo	L	M	J	Actividad	Evaluación
ABRIL	I. DISEÑO DE EXPERIMENTOS	5				
			7	8	Lab.0 - Intro a R	
		12				
			14	15	Lab.1 - ANOVA una vía	
		19				
			21	22	Lab.2 - Pruebas múltiples	Semana Universitaria
MAYO	II. REGRESIÓN	26				
			28	29	Lab.3 - Factoriales	Tarea No.1
		3				Feriado
			5	6	Lab.4 - Bloques	
		10			Clase de dudas	
			12	13		Examen 1
JUNIO	II. REGRESIÓN	17				
			19	20	Lab.5 - Regresión simple	
		24				Tarea No.2
			29	30	Lab.6 - Predicciones	
		31				
			6	7	Lab.7 - Diagnósticos	
JULIO	III. ANÁLISIS MULTIVARIADO	07				
			09	10	Lab.8 - Poisson - transformaciones	Anteproyecto
		14				
			16	17	Lab.9 - Logística	Tarea No.3
		21			Clase de dudas	
			23	24		Examen 2
JULIO	III. ANÁLISIS MULTIVARIADO	28				
			30	01	Lab.10 -Discriminante	
		05				
			07	08	Lab.11 - Componentes principales	
		12				
			14	15	Lab.12 - Cluster	Tarea No.4
JULIO	III. ANÁLISIS MULTIVARIADO	19			Clase de dudas	
			21	22		Examen 3
		26				
			28	29		Entrega y exposición del Trabajo final
		2				
			4	5		





8. Referencias bibliográficas

Catena et al. (2003). *Análisis multivariado: un manual para investigadores*. Madrid: Editorial Biblioteca Nueva.

Johnson, D. (2000). *Métodos multivariados aplicados al análisis de datos*. International Thomson Editores.

Montgomery, D. (2005). *Diseño y Análisis de Experimentos*. 2ª ed. Limusa Wiley.

Neter et al. (1996). *Applied Linear Statistical Models*. 3ª ed. WCB McGraw-Hill.

Ramsey, F.L. y D.W. Schafer (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis*. Duxbury, Australia: Thomson Learning.

Steel, R.G.D. y J.H. Torrie. (1985). *Bioestadística: principios y procedimientos*. Bogotá, Colombia: McGraw-Hill.

Zar, J.H. (1996). *Biostatistical Analysis*. 3a ed. Englewood Cliffs, N.J. : Prentice Hall.