



BIOESTADÍSTICA

AG100410

Código Interno ▲

2020

Primer Semestre

Doctorado en Ciencias Silvoagropecuarias y Veterinarias

CHRISTIAN HODAR QUIROGA	chodar@inta.uchile.cl	+562 29781532
IGOR PACHECO CRUZ	igor.pacheco@inta.uchile.cl	+562 29781552

Regular		15	5
		Máximo ▲	Mínimo ▲
		Cupos (N°)	
		Lunes ▶	
		Martes ▶	
		Miércoles ▶	
27 de marzo	17 de julio	Jueves ▶	
		Viernes ▶	09:30 – 13:30
Fecha de Inicio ▲	Fecha de término ▲	Día(s) ▲	Hora(s) ▲
Sin pre-requisitos		56	168
Pre-requisitos ▲		Directas ▲	indirectas ▲
		Número de horas (Totales) ▲	
		8	Créditos* ▲

*Sume horas (directas+Indirectas)/25. Coloque sólo valores enteros (Ej: 2,9=3; 2,4=2)

<p>Descripción y objetivos del curso</p> <p>El curso está orientado a revisar y poner en práctica de forma intensiva y práctica las metodologías de análisis exploratorio de datos e inferencia estadística, además de distintos métodos de construcción y verificación de modelos para el análisis avanzado de varianzas y cálculos de estimadores en datos de distribución no normal o desconocida, en datos obtenidos a través de experimentos diseñados según criterios objetivos. Presentará también una fuerte introducción a los métodos de análisis multivariado y descomposición de variables.</p> <p>El ramo además está contemplado para entregar conocimientos en la materia en un nivel medio/alto, considerando como aprobada la asistencia previa de los estudiantes a cursos de pregrado de estadística, y por ende con un conocimiento elemental de los elementos básicos acerca de estimadores, parámetros, probabilidad y distribuciones, los cuales sin embargo se revisarán en la primera clase. En el resto del curso, si bien se revisará la teoría, los elementos que subyacen a cada una de sus sesiones (ver programa), así como la utilidad de sus conceptos en el análisis de datos, pretenden profundizar el aspecto práctico del análisis, dedicando la mitad del tiempo a la práctica de análisis estadísticos utilizando el software de libre acceso R. En todas estas sesiones se utilizarán ejemplos obtenidos desde experimentos reales o desde datos simulados, aplicados a las distintas áreas del DCSAV: sistemas silvoagropecuarios, biomedicina, mejoramiento genético, conservación e inocuidad alimentaria.</p> <p>El carácter intensivo del curso será logrado a través de continuas evaluaciones clase a clase, que permitirán que el estudiante dé una continuidad al aprendizaje de las distintas materias contenidas en las 9 sesiones teóricas. Los alumnos además se deberán enfrentar al análisis de publicaciones de revistas de corriente principal, enfocados al análisis de los datos en ellas presentados.</p> <p>En conclusión, el curso pretende mostrar de manera sólida la teoría y las técnicas para la aplicación de los principales análisis estadísticos utilizados actualmente en trabajos publicados en revistas de corriente principal. Además de esto, los alumnos tendrán un conocimiento básico/medio del software R, útil para el análisis y representación gráfica avanzada de</p>
--



datos obtenidos experimentalmente provenientes de distintas disciplinas.

Objetivos:

Entregar conocimientos de teoría de los principales métodos de análisis estadístico, y su aplicación práctica mediante el uso del software R.

Metodología (Clases, seminarios, prácticos, otros)

- Las clases impartidas en cada módulo están diseñadas sobre impartir una base teórica, expuestas con material didáctico. Cada clase teórica será luego acompañada por ejercicios prácticos donde los contenidos expuestos sean aplicados mediante el uso del software R, plataforma especialmente construída para análisis estadístico. A partir de la sesión dos, en cada clase se efectuará un breve test (quiz) con respecto a los contenidos vistos en la clase anterior, con el fin de que el estudiante se mantenga activo con respecto a los contenidos entregados y sea capaz de generar un análisis reflexivo en torno a ellos. Adicionalmente, al final de cada sesión se les entregará a los alumnos un glosario de términos que le serán de utilidad para la clase siguiente.
- Se efectuarán dos sesiones de seminario en las cuales los alumnos deberán realizar un test introductorio de 3 preguntas acerca de los aspectos de análisis de datos y estadística aplicada empleados en 3 artículos que previamente serán entregados al curso. Por otro lado, los alumnos deberán exponer parte de estos artículos, enfocándose en el tipo de datos que se exponen, las preguntas que en ellos se plantean, y los análisis estadísticos que en ellos se lleva a cabo.
- Se efectuarán además trabajos prácticos, con fecha de entrega acotada, que corresponderán a tareas pendientes originadas en cada sesión práctica con el fin de que el alumno profundice los conocimientos prácticos adquiridos y adquiera familiaridad con el software R y sus módulos de análisis.
- Se efectuarán dos pruebas donde se evaluarán los conocimientos adquiridos durante el semestre.
- Ambos coordinadores estarán presentes en todas las sesiones, tanto teóricas como prácticas, al menos que existan razones de fuerza mayor que implique la ausencia de uno de ellos.
- La plataforma U-Cursos será el mecanismo oficial de coordinación en interacción de los coordinadores con los alumnos. Sin perjuicio de lo anterior, estos aceptan además recibir consultas y comentarios a través de sus correos institucionales.
- Los libros utilizados para la confección de este curso y nombrados en la bibliografía se encuentran completamente disponibles a través del portal web de la Biblioteca de la Universidad de Chile. El software R es de distribución gratuita y multiplataforma y se encuentra disponible en <https://www.r-project.org/>

Evaluación

ACTIVIDAD	%	Observaciones
Primera Prueba	30	* Los quiz corresponderán a pruebas simples de 3 preguntas, sobre temas tratados anteriormente y términos entregados en un glosario durante la sesión teórica anterior. Su objetivo es asegurar que el estudiante prepare los temas a tratarse en clase, mantenga activos los conocimientos entregados, desarrolle dudas para resolver en clases, y asegurar constancia en el estudio. Éstos se efectuarán los primeros 10 minutos de cada sesión teórica. Atrasos de menos de 10 minutos significarán menos tiempo para desarrollar el quiz. Atrasos de 10 o más minutos significarán no rendición del quiz y por ende, nota 1.
Segunda Prueba	30	
Promedio quiz (n = 8)*	15	
Promedio trabajos prácticos (n=8)**	15	
Seminario***	10	
TOTAL	100	



	<p>** Los trabajos prácticos corresponderán a tareas pendientes en cada sesión práctica, y que se deberán entregar en un plazo máximo de 48 horas después de terminada ésta última. Atrasos de 48 horas adicionales significarán una nota máxima de 4. Atrasos superiores a 96 horas, significarán nota 1. Se hará entrega a los alumnos de las preguntas que ellos responderán utilizando los análisis aprendidos en clase. Los profesores estarán disponibles sólo las primeras 48 horas después de la clase vía plataforma u-cursos.</p> <p>*** Se efectuará una sesión de seminario en las cuales los alumnos deberán exponer parte de artículos elegidos por ellos, enfocándose en el tipo de datos que se presentan, las preguntas que en ellos se plantean, y los análisis estadísticos que en ellos se lleva a cabo. Las actividades de seminario tendrán cada una un valor del 5% de la nota del curso, y se evalúa por presentación del problema, planteamiento del proceso estadístico y participación en los trabajos de los demás compañeros.</p>
--	--

Profesores participantes

Nombres y Grados Académicos	Categoría Académica	Institución	Participación*
Igor Pacheco Cruz, Ing. Biotecnología Molec. PhD.	Profesor Asistente	Universidad de Chile	Profesor responsable
Christian Hodar, Ing. Civil Biotecnología. PhD.	Profesor Asistente	Universidad de Chile	Profesor responsable
José Manuel Yáñez, Medico Veterinario, PhD	Profesor Asociado	Universidad de Chile	Profesor colaborador

***Profesor Responsable:** Formalmente encargado del curso y tiene la atribución de firmar el acta de evaluación de los estudiantes.

Colaborador: Integrante del equipo docente del curso, que realiza actividades de apoyo, fundamentales o complementarias para la realización del curso, y cuya participación tiene una duración mayor a dos semanas. Ejemplos de este nivel de participación son: profesor a cargo de trabajos prácticos, profesor que dicta las clases teóricas de un (o más de un) capítulo o módulo del programa, profesor encargado de alguna actividad específica complementaria.

Invitado: corresponde a un profesor que dicta entre una y cuatro clases de un curso, o que participa en una actividad específica complementaria.

Ayudante: corresponde a una participación de apoyo al profesor responsable en sesiones de ayudantía, evaluaciones, preparación de material de apoyo y/o apoyo en laboratorios, trabajos prácticos y talleres.



Contenidos

Fecha	Contenidos	Profesor	Número de horas	
			Directas	Indirectas
27/03/2020	Introducción al curso. Descripción del programa y de los contenidos. Bibliografía. Software R. Entrega del glosario para la próxima clase.	Igor Pacheco / Christian Hodar	4	12
03/04/2019	<p>QUIZ 1. Preguntas sobre conceptos del glosario enviado la clase anterior.</p> <p>SECCIÓN TEÓRICA 1a. Mención de conceptos base para el curso (pt1).</p> <ul style="list-style-type: none"> ● Probabilidad y estimadores. ● Distribuciones de probabilidad (Normal, Binomial, Poisson, X^2, etc.). ● Inferencias y pruebas de hipótesis (test t student, test de X^2, test de Fisher, test de Wilcoxon-Mann-Whitney). ● Entrega de glosario para la próxima clase. <p>SESIÓN PRÁCTICA 1a. Introducción al análisis estadístico utilizando el software R (pt1)</p> <ul style="list-style-type: none"> ● Aspectos básicos de uso de R (operaciones básicas, variables, tipos de datos, vectores, funciones y documentación, datos nulos; estadísticas descriptivas). ● Estructuras de datos típicas en R y sus diferencias (data frames, listas, matrices y arrays). ● Lectura de datos en R (desde: csv, xls, bases de datos, otros softwares de estadística, etc). ● Visualización de datos en R- paquete “ggplot” y “lattice” (parte 1). ● Trabajo práctico 1a. 	Igor Pacheco / Christian Hodar	4	12
10/04/2020	Viernes santo			
17/04/2020	<p>QUIZ 2. Preguntas sobre conceptos del glosario estudiado y de la clase anterior.</p> <p>SECCIÓN TEÓRICA 1b. Mención de conceptos base para el curso.</p> <ul style="list-style-type: none"> ● Probabilidad y estimadores. ● Distribuciones de probabilidad (Normal, Binomial, Poisson, X^2, etc.). ● Inferencias y pruebas de hipótesis (test t student, test de X^2, test de Fisher, test de Wilcoxon-Mann-Whitney). ● Diseño experimental (Tres principios básicos, Principio de Pareto de importancia relativa de los efectos) ● Entrega de glosario para la próxima clase. <p>SECCIÓN PRÁCTICA 1b. Introducción al análisis estadístico utilizando el software R (pt2)</p> <ul style="list-style-type: none"> ● Aspectos básicos de uso de R (operaciones básicas, variables, tipos de datos, vectores, funciones y 	Igor Pacheco / Christian Hodar	4	12



	<p>documentación, datos nulos; estadísticas descriptivas).</p> <ul style="list-style-type: none"> ● Estructuras de datos típicas en R y sus diferencias (data frames, listas, matrices y arrays). ● Lectura de datos en R (desde: csv, xls, bases de datos, otros softwares de estadística, etc). ● Visualización de datos en R- paquete “ggplot” y “lattice” (parte 2). ● Trabajo práctico 1b. 			
24/03/2020	<p>QUIZ 3. Preguntas sobre conceptos del glosario estudiado y de la clase anterior.</p> <p>SECCIÓN TEÓRICA 1c. Mención de conceptos base para el curso.</p> <ul style="list-style-type: none"> ● Probabilidad y estimadores. ● Distribuciones de probabilidad (Normal, Binomial, Poisson, X^2, etc.). ● Inferencias y pruebas de hipótesis (test t student, test de X^2, test de Fisher, test de Wilcoxon-Mann-Whitney). ● Diseño experimental (Tres principios básicos, Principio de Pareto de importancia relativa de los efectos) ● Entrega de glosario para la próxima clase. <p>SECCIÓN PRÁCTICA 1c. Introducción al análisis estadístico utilizando el software R (pt2)</p> <ul style="list-style-type: none"> ● Aspectos básicos de uso de R (operaciones básicas, variables, tipos de datos, vectores, funciones y documentación, datos nulos; estadísticas descriptivas). ● Estructuras de datos típicas en R y sus diferencias (data frames, listas, matrices y arrays). ● Lectura de datos en R (desde: csv, xls, bases de datos, otros softwares de estadística, etc). ● Visualización de datos en R- paquete “ggplot” y “lattice” (parte 2). ● Trabajo práctico 1b. 	Igor Pacheco / Christian Hodar	4	12
01/05/2020	Dial del trabajador			
08/05/2020	<p>QUIZ 4. Preguntas sobre conceptos del glosario estudiado y de la clase anterior.</p> <p>SESIÓN TEÓRICA 2. Pruebas de comparación de medias muestrales: paramétricas y no paramétricas.</p> <ul style="list-style-type: none"> ● Pruebas paramétricas: test z, test t student ● Pruebas no paramétricas: chi cuadrado, Wilcoxon, U-Mann-Whitney, Kolmogorov-Smirnov. <p>SESIÓN PRÁCTICA 2. Pruebas de comparación de medias muestrales: no paramétricas.</p> <ul style="list-style-type: none"> ● Pruebas no paramétricas: chi cuadrado, Wilcoxon, U-Mann-Whitney, Kolmogorov-Smirnov. 	Igor Pacheco / Christian Hodar	4	12



15/05/2020	<p>SESIÓN TEÓRICA 3. Regresión Lineal</p> <ul style="list-style-type: none"> ● Correlación y Regresión Lineal (coeficientes de Pearson y Spearman) <p>SESIÓN PRÁCTICA 3.</p> <ul style="list-style-type: none"> ● funciones, revisar datos, iteraciones, grupos, manipulación y formateo de datos ● Revisión de las funciones $lm()$ ● Aplicación de test de correlación, regresión lineal y logística, utilizando ejemplos de datos simulados en R, disponibles en la web. ● Trabajo práctico 2. 	Igor Pacheco / Christian Hodar	4	12
22/05/2020	Seminario 1.	Igor Pacheco / Christian Hodar	4	12
29/05/2020	Prueba 1 (30%)	Igor Pacheco / Christian Hodar	4	12
05/06/2020	<p>QUIZ 5. Preguntas sobre conceptos del glosario estudiado y de la clase anterior.</p> <p>SESIÓN TEÓRICA 4. Análisis de Varianza (ANOVA)</p> <ul style="list-style-type: none"> ● Análisis de Varianza a una vía (principios subyacentes, técnica y conclusiones; supuestos para ANOVA; tests alternativos no paramétricos: Kruskal-Wallis, etc) ● Análisis de Varianza a Múltiples Vías (principios subyacentes, técnica y conclusiones; supuestos para ANOVA; tests alternativos) ● Análisis de Covarianza (ANCOVA), principales métodos. <p>SESIÓN PRÁCTICA 4.</p> <ul style="list-style-type: none"> ● Aplicación de los test aprendidos con ejemplos de sets de datos simulados en R y disponibles. ● Función $aov()$ ● Trabajo práctico 3. 	Igor Pacheco / Christian Hodar	4	12
12/06/2020	<p>QUIZ 6. Preguntas sobre conceptos del glosario estudiado y de la clase anterior.</p> <p>SESIÓN TEÓRICA 5. Introducción a los Modelos lineales generalizados mixtos (GLMM): análisis para variables continuas y discretas.</p> <ul style="list-style-type: none"> ● Métodos de máxima verosimilitud, desviación como parámetro de bondad de ajuste. ● Definición de GLM ● Componentes de los GLM: estocásticos y sistemáticos ● Ajustes de modelos lineales generalizados: ejemplos básicos <p>SESIÓN PRÁCTICA 5.</p>	Igor Pacheco / Christian Hodar	4	12



	<ul style="list-style-type: none"> ● Uso del paquete lme4 (parte 1) ● Aplicación de modelos lineales generalizados a ejemplos de sets de datos disponibles en internet. ● Visualización de resultados. ● Trabajo práctico 4 			
19/06/2020	<p>QUIZ 7. Preguntas sobre conceptos del glosario estudiado y de la clase anterior.</p> <p>SESIÓN TEÓRICA 6. Descomposición de matrices y reducción de dimensionalidad.</p> <ul style="list-style-type: none"> ● Descomposición de matrices a factores. ● Análisis de componentes principales (PCA). ● Análisis factorial (FA). ● Escalamiento multidimensional (MDS). <p>Entrega de glosario para la próxima clase</p> <p>SESIÓN PRÁCTICA 6. Ejemplos prácticos de análisis dimensional usando R</p> <ul style="list-style-type: none"> ● Visualización de datos multivariados ● Manipulación de matrices de y determinación de factores. ● Inspección gráfica de dimensiones. ● Ejercicios de PCA, FA y MDS ● Trabajo práctico 6 	Igor Pacheco / Christian Hodar	4	12
26/06/2020	<p>QUIZ 8. Preguntas sobre conceptos del glosario estudiado y de la clase anterior.</p> <p>SECCIÓN TEÓRICA 7. Clasificación y análisis de agrupaciones.</p> <ul style="list-style-type: none"> ● Agrupaciones jerárquicas (Clusters) ● Agrupaciones no supervisadas (K-means) ● Agrupaciones difusas (Fuzzy means) ● Análisis discriminante (DA). <p>SECCIÓN PRÁCTICA 7. Ejemplos prácticos de clasificación usando R</p> <ul style="list-style-type: none"> ● Visualización de grupos pre y post clasificación ● Ejercicios de agrupación, usando diferentes algoritmos de clasificación. ● Ejercicios de análisis discriminante. ● Trabajo práctico 7 	Christian Hodar	4	12
03/07/2020	INTRODUCCIÓN A LA ESTADISTICA BAYESIANA	José Manuel Yáñez	4	12
10/07/2020	SEMINARIO 2	Igor Pacheco / Christian Hodar/José Manuel Yáñez	4	12



17/07/2020	Prueba 2 (30%)	Igor Pacheco / Christian Hodar	4	12
		Total	60	180

Bibliografía

1. Islam and Al-Shiha, 2018. Foundations of Biostatistics. Springer Nature Singapore
2. Schumacker, 2013. Understanding Statistics Using R. Springer.
3. Anon, 2018. Generalized Linear Models with Examples In R. Springer Science+business Media.
4. Galecki, A., 2013. Linear Mixed-effects Models Using R: A Step-by-step Approach (springer Texts In Statistics). Springer.
5. Sahu, 2016. Applied Statistics for Agriculture, Veterinary, Fishery, Dairy and Allied Fields. Springer India.
6. Everitt and Hothorn, 2011. An Introduction to Applied Multivariate Analysis with R. Springer
7. Härdle and Simar, 2015. Applied Multivariate Statistical Analysis. Springer