



**BIOESTADÍSTICA**

**AG100410**

Código Interno ▲

**2019**

**Primer Semestre**

**Doctorado en Ciencias Silvoagropecuarias y Veterinarias**

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

<b>Regular</b>		<b>15</b>	<b>5</b>
		<b>Máximo ▲</b>	<b>Mínimo ▲</b>
		<b>Cupos (N°)</b>	
		<b>Lunes ►</b>	
		<b>Martes ►</b>	
		<b>Miércoles ►</b>	
<b>11 de abril</b>	<b>14 de junio</b>	<b>Jueves ►</b>	
		<b>Viernes ►</b>	<b>09:30 – 13:30</b>
<b>Fecha de Inicio ▲</b>	<b>Fecha de término ▲</b>	<b>Día(s) ▲</b>	<b>Hora(s) ▲</b>
<b>Sin pre-requisitos</b>		<b>56</b>	<b>168</b>
<b>Pre-requisitos ▲</b>		<b>Directas ▲</b>	<b>indirectas ▲</b>
		<b>Número de horas (Totales) ▲</b>	
			<b>8</b>
			<b>Créditos* ▲</b>

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

<b>Descripción y objetivos del curso</b>
<p>El curso está orientado a revisar y poner en práctica de forma intensiva 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.

**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 construida 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 = 10)*	15	
Promedio trabajos prácticos (n=8)**	15	
Seminarios***	10	
<b>TOTAL</b>	100	** 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



	<p>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 efectuarán dos sesiones de seminario en las cuales i) se realizará un Quiz de 3 preguntas acerca de los aspectos de análisis de datos y estadística aplicada, a los datos que se muestran en un grupo de 3 artículos científicos previamente entregados al curso; ii) 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. Las pruebas de seminario tendrán cada una un valor del 5% de la nota del curso. Los atrasos funcionan tal y como en los Quiz de sesiones teóricas.</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

**\*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
<b>MARZO</b>				
15/03/2019	<b>Introducción al curso. Descripción del programa y de los contenidos. Bibliografía. Software R. Entrega del glosario para la próxima clase.</b>	Igor Pacheco / Christian Hodar	4	12
29/04/2019	<p><b>QUIZ 1. Preguntas sobre conceptos del glosario estudiado y de la clase anterior.</b></p> <p><b>SECCIÓN TEÓRICA 1a. Mención de conceptos base para el curso (pt1).</b></p> <ul style="list-style-type: none"> <li>● Probabilidad y estimadores.</li> <li>● Distribuciones de probabilidad (Normal, Binomial, Poisson, <math>X^2</math>, etc.).</li> <li>● Inferencias y pruebas de hipótesis (test t student, test de <math>X^2</math>, test de Fisher, test de Wilcoxon-Mann-Whitney).</li> <li>● Diseño experimental (Tres principios básicos, Principio de Pareto de importancia relativa de los efectos)</li> <li>● Entrega de glosario para la próxima clase.</li> </ul> <p><b>SESIÓN PRÁCTICA 1a. Introducción al análisis estadístico utilizando el software R (pt1)</b></p> <ul style="list-style-type: none"> <li>● Aspectos básicos de uso de R (operaciones básicas, variables, tipos de datos, vectores, funciones y documentación, datos nulos; estadísticas descriptivas).</li> <li>● Estructuras de datos típicas en R y sus diferencias (data frames, listas, matrices y arrays).</li> <li>● Lectura de datos en R (desde: csv, xls, bases de datos, otros softwares de estadística, etc).</li> <li>● Visualización de datos en R- paquete “ggplot” y “lattice” (parte 1).</li> <li>● Trabajo práctico 1a.</li> </ul>	Igor Pacheco / Christian Hodar	4	12
<b>ABRIL</b>				
05/04/2019	<p><b>QUIZ 2. Preguntas sobre conceptos del glosario estudiado y de la clase anterior.</b></p> <p><b>SECCIÓN TEÓRICA 1b. Mención de conceptos base para el curso.</b></p> <ul style="list-style-type: none"> <li>● Probabilidad y estimadores.</li> <li>● Distribuciones de probabilidad (Normal, Binomial, Poisson, <math>X^2</math>, etc.).</li> <li>● Inferencias y pruebas de hipótesis (test t student, test de <math>X^2</math>, test de Fisher, test de Wilcoxon-Mann-Whitney).</li> <li>● Diseño experimental (Tres principios básicos, Principio de Pareto de importancia relativa de los</li> </ul>	Igor Pacheco / Christian Hodar	4	12



	<p>efectos)</p> <ul style="list-style-type: none"> <li>Entrega de glosario para la próxima clase.</li> </ul> <p><b>SECCIÓN PRÁCTICA 1b. Introducción al análisis estadístico utilizando el software R (pt2)</b></p> <ul style="list-style-type: none"> <li>Aspectos básicos de uso de R (operaciones básicas, variables, tipos de datos, vectores, funciones y documentación, datos nulos; estadísticas descriptivas).</li> <li>Estructuras de datos típicas en R y sus diferencias (data frames, listas, matrices y arrays).</li> <li>Lectura de datos en R (desde: csv, xls, bases de datos, otros software de estadística, etc).</li> <li>Visualización de datos en R- paquete “ggplot” y “lattice” (parte 2).</li> <li>Trabajo práctico 1b.</li> </ul>			
12/04/2019	<p><b>QUIZ 3. Preguntas sobre conceptos del glosario estudiado y de la clase anterior.</b></p> <p><b>SESIÓN TEÓRICA 2. Regresión.</b></p> <ul style="list-style-type: none"> <li>Correlación y Regresión Lineal (coeficientes de Pearson y Spearman: definiciones y propiedades; análisis de regresión, componentes, sus propiedades y pruebas de hipótesis en base a la pendiente y el intercepto de una función de regresión)</li> <li>Análisis de respuestas categóricas y regresión logística</li> <li>Entrega de glosario para la próxima clase.</li> </ul> <p><b>SESIÓN PRÁCTICA 2.</b></p> <ul style="list-style-type: none"> <li>funciones, revisar datos, iteraciones, grupos, manipulación y formateo de datos</li> <li>Revisión de las funciones <code>lm()</code> y <code>gls()</code></li> <li>Visualización de datos en R- paquete “ggplot” y “lattice” (parte 3).</li> <li>Aplicación de test de correlación, regresión lineal y logística, utilizando ejemplos de datos simulados en R, disponibles en la web.</li> <li>Trabajo práctico 2.</li> </ul>	Igor Pacheco / Christian Hodar	4	12
16/04/2019	<p><b>QUIZ 4. Preguntas sobre conceptos del glosario estudiado y de la clase anterior.</b></p> <p><b>SESIÓN TEÓRICA 3. Análisis de Varianza (ANOVA)</b></p> <ul style="list-style-type: none"> <li>Análisis de Varianza a una vía (principios subyacentes, técnica y conclusiones; supuestos para ANOVA; tests alternativos no paramétricos: Kruskal-Wallis, etc)</li> <li>Análisis de Varianza a Múltiples Vías (principios subyacentes, técnica y conclusiones; supuestos para ANOVA; tests alternativos)</li> <li>Análisis de Covarianza (ANCOVA), principales</li> </ul>	Igor Pacheco / Christian Hodar	4	12



	<p>métodos.</p> <p><b>SESIÓN PRÁCTICA 3.</b></p> <ul style="list-style-type: none"> <li>• Aplicación de los test aprendidos con ejemplos de sets de datos simulados en R y disponibles.</li> <li>• Función <code>aov()</code></li> <li>• Trabajo práctico 3.</li> </ul>			
26/04/2019	<p><b>QUIZ 5. Preguntas sobre conceptos del glosario estudiado y de la clase anterior.</b></p> <p><b>SESIÓN TEÓRICA 4. Modelos lineales generalizados (GLM): análisis para variables binarias.</b></p> <ul style="list-style-type: none"> <li>• Métodos de máxima verosimilitud, desviación como parámetro de bondad de ajuste.</li> <li>• Definición de GLM</li> <li>• Componentes de los GLM: estocásticos y sistemáticos</li> <li>• Ajustes de modelos lineales generalizados: ejemplos básicos</li> </ul> <p><b>SESIÓN PRÁCTICA 4.</b></p> <ul style="list-style-type: none"> <li>• Uso del paquete <code>lme4</code> (parte 1)</li> <li>• Aplicación de modelos lineales generalizados a ejemplos de sets de datos disponibles en internet.</li> <li>• Visualización de resultados.</li> <li>• Trabajo práctico 4</li> <li>•</li> </ul>	Igor Pacheco / Christian Hodar	4	12
<b>MAYO</b>				
03/05/2019	<p><b>QUIZ S1.</b></p> <p><b>Seminario 1.</b></p>	Igor Pacheco / Christian Hodar	4	12
10/05/2019	<p><b>Prueba 1 (30%)</b></p>	Igor Pacheco / Christian Hodar	4	12
17/05/2019	<p><b>QUIZ 6. Preguntas sobre conceptos del glosario estudiado y de la clase anterior.</b></p> <p><b>SESIÓN TEÓRICA 5. Modelos lineales mixtos.</b></p> <ul style="list-style-type: none"> <li>• Modelos lineales con varianza homogénea</li> <li>• Modelos lineales con varianza heterogénea</li> <li>• Modelos lineales con factores fijos y errores correlacionados</li> <li>• Ajustes de modelos lineales mixtos: ejemplos básicos</li> </ul> <p><b>SESIÓN PRÁCTICA 5.</b></p> <ul style="list-style-type: none"> <li>• Uso del paquete <code>lme4</code> (parte 2)</li> <li>• Aplicación de modelos lineales mixtos a ejemplos de sets de datos disponibles en internet; funciones <code>lme()</code> y <code>lmer()</code></li> <li>• Visualización de resultados.</li> <li>• Trabajo práctico 5</li> </ul>	Igor Pacheco / Christian Hodar	4	12



24/05/2019	<p><b>QUIZ 7. Preguntas sobre conceptos del glosario estudiado y de la clase anterior.</b></p> <p><b>SECCIÓN TEÓRICA 6. Descomposición de matrices y reducción de dimensionalidad.</b></p> <ul style="list-style-type: none"> <li>• Descomposición de matrices a factores.</li> <li>• Análisis de componentes principales (PCA).</li> <li>• Análisis factorial (FA).</li> <li>• Escalamiento multidimensional (MDS).</li> <li>• Entrega de glosario para la próxima clase.</li> </ul> <p><b>SECCIÓN PRÁCTICA 6. Ejemplos prácticos de análisis dimensional usando R</b></p> <ul style="list-style-type: none"> <li>• Visualización de datos multivariados</li> <li>• Manipulación de matrices de y determinación de factores.</li> <li>• Inspección gráfica de dimensiones.</li> <li>• Ejercicios de PCA, FA y MDS</li> <li>• Trabajo práctico 6</li> </ul>	Igor Pacheco / Christian Hodar	4	12
<b>JUNIO</b>				
31/05/2019	<p><b>QUIZ 8. Preguntas sobre conceptos del glosario estudiado y de la clase anterior.</b></p> <p><b>SECCIÓN TEÓRICA 7. Clasificación y análisis de agrupaciones.</b></p> <ul style="list-style-type: none"> <li>• Agrupaciones jerárquicas (Clusters)</li> <li>• Agrupaciones no supervisadas (K-means)</li> <li>• Agrupaciones difusas (Fuzzy means)</li> <li>• Análisis discriminante (DA).</li> </ul> <p><b>SECCIÓN PRÁCTICA 7. Ejemplos prácticos de clasificación usando R</b></p> <ul style="list-style-type: none"> <li>• Visualización de grupos pre y post clasificación</li> <li>• Ejercicios de agrupación, usando diferentes algoritmos de clasificación.</li> <li>• Ejercicios de análisis discriminante.</li> <li>• Trabajo práctico 7</li> </ul>	Igor Pacheco / Christian Hodar	4	12
07/06/2019	<b>QUIZ S2 Seminario 2</b>	Igor Pacheco / Christian Hodar	4	12
14/06/2019	<b>Prueba 2 (30%)</b>	Igor Pacheco / Christian Hodar	4	12
<b>Total</b>			<b>48</b>	<b>144</b>

## 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