



UNIVERSIDAD NACIONAL DE LA PLATA
FACULTAD DE CIENCIAS AGRARIAS Y FORESTALES
PROSECRETARÍA DE POSGRADO
Avda. 60 y 119 – La Plata – C.P. (1900) – C.C. 31
<http://www.agro.unlp.edu.ar>

Curso de Posgrado:

“Introducción al Análisis Estadístico Robusto con R”

Curso Acreditado a Carreras de Posgrado Especialización, Maestrías y Doctorado (Artículo 3 de la Ordenanza CS N°261/19)

Docentes: Dr. Luis Acuña Rello – Dra. Eleana Spavento

Carga Horaria Total: 75 horas

Fecha de dictado: desde el **2 de noviembre** al **14 de diciembre de 2020**

1.1. Fundamentación de la Propuesta

La estadística constituye una herramienta fundamental para la toma de decisiones, el análisis crítico de resultados y la obtención de conclusiones, constituyendo un nexo común en la mayoría de las investigaciones científicas; nos aporta herramientas que van, desde el análisis e interpretación de datos (estadística descriptiva) hasta el proceso de predicción y toma de decisiones (estadística inferencial). Actualmente, es tal la importancia que ha adquirido que aparece en prácticamente todas las áreas de trabajo, siendo un factor clave para la toma de decisiones a partir de datos observados. Sin embargo, para ello resulta imprescindible que el profesional/investigador cuente con los conocimientos básicos sobre estadística para no llegar a conclusiones equivocadas.

Desde hace algunos años, y en ocasiones apoyándose en la gran mejora de la capacidad de cálculo de los ordenadores, junto a los métodos clásicos estadísticos, se han ido desarrollando técnicas capaces de dar buenos resultados sin tener que ceñirse a las exigencias formales que impone la estadística lineal.

En este contexto, el software de distribución libre y gratuito (Licencia GNU) “R” se ha convertido en una herramienta ampliamente utilizada en la investigación, tanto desde el punto de vista estadístico básico como aplicado. Es un sistema derivado del lenguaje S que

se utiliza para análisis estadísticos y producción de gráficos, con una doble naturaleza: programa de análisis y lenguaje de programación. Se caracteriza por su versatilidad, por la participación constante de una comunidad de investigadores en su desarrollo que permiten que existan un serie de ayudas y foros que mejoran/aceleran el proceso de aprendizaje, por la mayor posibilidad de funciones con respecto a otros software, por la posibilidad que le brinda a cada usuario a crear sus funciones combinando/aplicando funciones existentes, por las múltiples aplicaciones para la generación de gráficos, entre otras consideraciones que hacen que en la actualidad se puedan visualizar un número creciente de trabajos que citan este software, como así también, la presencia de rutinas en este lenguaje publicadas como material suplementarios en revistas internacionales. De todos modos, el hecho de no disponer de un interfaz basado en menús hace que el proceso de aprendizaje sea más costoso respecto a otros paquetes comerciales (SPSS, STATA, SAS), pero una vez superado este aprendizaje, R se convierte en una herramienta ideal para el análisis estadístico, programación de funciones y para la representación gráfica de resultados, consideraciones que justifican su enseñanza en el ámbito universitario-científico. Es por ello que mediante esta actividad de postgrado, se pretende que los participantes, profesionales del ámbito académico-científico-independiente, se familiaricen en el entorno de trabajo de R con el fin de realizar el tratamiento, análisis de datos, predicciones entre otras consideraciones detalladas dentro del programa de actividad propuesto. Para ello, si bien se plantea este curso de carácter fundamentalmente aplicado, el mismo será complementado con aquellos elementos teóricos necesarios para elaborar estrategias para la toma, análisis e interpretación de datos y para la selección de las técnicas estadísticas apropiadas para cada caso. Esto se logrará a través de distintas actividades que le permitirán al alumno adquirir habilidades y destrezas para la formulación, análisis y resolución de problemas típicos de las áreas de incumbencia.

1.4. Objetivos:

- Introducir al alumno en el uso del lenguaje y manejo del software R.
- Revisar las técnicas clásicas fundamentales de análisis estadístico.
- Introducir al alumno en las técnicas de análisis robusto.

1.5. Contenidos y Bibliografía

Programa: Contenidos

Unidad 1: Introducción al lenguaje y manejo de R

Tema 1.1- Instalando R: aspectos más importantes del lenguaje estadístico R; instalación y primeros acercamientos al uso de R.

Tema 1.2- Ayuda en R. Preparación e importación de datos: inicio de la actividad práctica a través de la carga de datos obtenidos de una actividad previamente preestablecida y entregados por el docente a cargo de la actividad.

Unidad 2: Análisis estadístico descriptivo

Tema 2.1- Análisis descriptivo: interpretación de parámetros de centralidad, dispersión, de forma, entre otros. Se trabajará con los datos previamente cargados a fin de realizar una descripción estadística de la población.

Tema 2.2- Pruebas de normalidad: interpretación de parámetros. A partir de los datos poblacionales inicialmente cargados se realizará una evaluación sobre el supuesto de normalidad.

Tema 2.3- Pruebas de igualdad de varianza (homocedasticidad): interpretación de parámetros. Se procederá al igual que lo detallado en el tema 2.1 y 2.1 para la evaluación del supuesto de homocedasticidad.

Unidad 3: Análisis estadístico inferencial

Tema 3.1- Comparaciones: de dos muestras: a partir de los datos cargados inicialmente se realizará una comparación entre las situaciones que se planteen en la etapa inicial de la actividad.

Tema 3.2- Comparaciones: múltiples: análisis e interpretación de métodos. Se procederá a partir de los datos y situaciones planteadas en la etapa inicial de la actividad.

Tema 3.3- Predicción. Modelo lineal: a partir de los datos cargados y de acuerdo con la situación que se plantee en la etapa inicial de la actividad, se llevará a cabo la predicción de la variable dependiente a partir de variable independiente (predictora) previamente establecidas.

Tema 3.4- Predicción. Regresiones "linealizables": a partir de un ejemplo general se atenderá al tratamiento estadístico de relaciones no lineales, pero fácilmente linealizables.

Unidad 4: Técnicas de análisis robusto

Tema 4.1- Predicción: regresiones robustas. A partir de los datos inicialmente cargados, se obtendrán diferentes modelos de regresión que utilizan estimaciones robustas frente a la influencia distorsionadora de las observaciones influyentes o aberrantes.

Tema 4.2- Predicción: regresión lineal múltiple y múltiple robusta. Trabajando con los datos inicialmente cargados se obtendrán modelos lineales y modelos robustos con dos o más variables predictoras.

Tema 4.3- Modelo lineal generalizado (GLM, de *Generalized Linear Models*). Con los datos existentes se realizará un GLM, esto es, un análisis de tipo regresión en el que los errores (residuos) se distribuyan según distribuciones distintas de la normal.

Tema 4.4- Métodos de remuestreo. Se introducirá a los participantes en las técnicas básicas de remuestreo (permutaciones, jackknife, bootstrap y validación cruzada) utilizando el paquete R.

Tema 4.5- Redes neuronales en R. Se introducirá en la utilización de las redes neuronales implementadas en los paquetes de R como herramienta de predicción y clasificación.

Programa: Bibliografía

Arriaza Gómez, A.J.; Fernández Palacín, F.; López Sánchez, M.A.; Muñoz Márquez, M.; Pérez Plaza, S.; Sánchez Navas, A. 2008. Estadística Básica con R y R-commander. Servicio de Publicaciones de la Universidad de Cádiz. 160 pp.

Canty, A.; Davison, A.; Hinkley, D.; Ventura, V. 2006. Bootstrap diagnostics and remedies. *The Canadian Journal of Statistics*. Vol. 34, No. 1, 2006. Pages 5-27.

Chang, W. 2013. R Graphics Cookbook. Published by O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472.

Chernick, M.; LaBudde, R. 2011. An introduction to bootstrap methods with applications to R. John Wiley & Sons, Inc. ISBN 978-0-470-46704-6

Coghlan, A. 2014. A Little Book of R For Multivariate Analysis. 51 pp.

Contreras García, J.M.; Molina Portillo, E.; Arteaga Cezón, P. 2010. Introducción a la programación estadística con R para profesores. ISBN: 978-84-693-4859-8.

Crawley, M. 2005. Statistics, An Introduction Using R. John Wiley & Sons, Inc. ISBN: 978111894109-6.

Dalgaard, P. 2008. Introductory Statistics with R. Springer. Second Edition. DOI: 10.1007/978-0-387-79054-1. 370 pp.

Faraway, J. 2016. Extending the Linear Model with R. Generalized Linear, Mixed Effects and Nonparametric Regression Models. Chapman & Hall/CRC Taylor & Francis Group. 312 pp.

García-Pérez, A. 2005. Métodos Avanzados de Estadística Aplicada. Técnicas Avanzadas. Editorial UNED. Colección Educación Permanente. 255 pp.

Gschwandtner, M.; Filzmoser, P. 2012. Computing Robust Regression Estimators: Developments since Dutter (1977). Austrian Journal of Statistics. Vol 41:1. Pages 45–58

Hastie, T.; Tibshirani, R.; Friedman, J. 2008. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer Series in Statistics

James, G.; Witten, D.; Hastie, T.; Tibshirani, R. 2013. An Introduction to Statistical Learning with Applications in R. Springer Texts in Statistics. DOI 10.1007/978-1-4614-7138-7. 434 pp.

Kabacoff, R. 2011. R In Action. Data analysis and graphics with R. Manning Publications Co. ISBN: 9781935182399.

Kuhnert, P.; Venables, B. 2005. An Introduction to R: Software for Statistical Modelling Computing. CSIRO Mathematical and Information Sciences, Cleveland, Australia.

Large, B. 2014. R Bootstrap Examples. Chapter: 3. 11 pp.

Maindonald, J. H. 2008. Using R for Data Analysis and Graphics. Introduction, Code and Commentary. © J. H. Maindonald 2000, 2004, 2008.

Mair, P.; Wilcox, R. 2016. Robust Statistical Methods in R Using the WRS2 Package. Journal of Statistical Software. Vol VV:II. doi: 10.18637/jss.v000.i00. 32 pp.

Mangiafico, S. 2016. Summary and analysis of extension program evaluation in R. Rutgers Cooperative Extension, New Brunswick, NJ. Web: rcompanion.org/handbook/

McDonald, J. 2014. Handbook of biological statistics. Sparky House Publishing Baltimore, Maryland, U.S.A.

Quick, J. 2010. Statistical Analysis with R. Packt Publishing, Birmingham-Mumbai. 300 pp.

Wilcox, R.; Rousselet, G. 2018. A guide to robust statistical methods in neuroscience. Current Protocols in Neuroscience. <https://doi.org/10.1002/cpns.41>. 30 pp.

Wilcox, R.; Keselman, H. 2003. Modern Robust Data Analysis Methods: Measures of Central Tendency. Psychological Methods. Vol 8:3. Pages 254-274. DOI: 10.1037/1082-989X.8.3.254. 21 pp.

Programa: Páginas web de relevancia

<http://cran.r-project.org>

<https://www.statmethods.net/management/subset.html>

<https://rveryday.wordpress.com/2016/11/29/5-ways-to-subset-a-data-frame-in-r/>

<https://cran.r-project.org/doc/contrib/grafi3.pdf>

<http://www.statmethods.net/advgraphs/parameters.html>

<http://www.elsevier.es/es-revista-atencion-primaria-27-articulo-utilizacion-metodos-robustos-estadistica-inferencial-13049898>

<https://stats.idre.ucla.edu/r/dae/robust-regression/>

http://www.uam.es/personal_pdi/ciencias/jspinill/CFCUAM2014/MetodosAutodocimantes-CFCUAM2014.html

1.6. Metodología

- Clases asincrónicas con tutorías de lectura y/o resolución de ejercitación práctica.
- Clases a distancia sincrónicas obligatorias con aportes teóricos y resoluciones prácticas.
- Tutorías de lectura-resolución de ejercicios prácticos y realización de informe final a distancia (a requerimiento).

1.7. Evaluación:

Las condiciones de aprobación del curso contemplan el 80% de asistencia a las clases sincrónicas obligatorias (modalidad a distancia), la realización y entrega digital de actividades parciales y la realización y entrega digital de un informe final teórico-práctico, individual, que incorporará los conocimientos y las resoluciones de problemas constituidos durante el desarrollo del curso y tutorías a distancia. La extensión del mismo y el plazo de entrega serán indicados durante el desarrollo del curso.

1.8. Cupo de alumnos para el dictado (Mínimo y máximo).

Número de alumnos mínimo: 10

Número de alumnos máximo: 25

Destinado a: Ingenieros Forestales, Ingenieros Agrónomos, Ingenieros en Recursos Naturales, Ecólogos y carreras afines.

Requisitos: conocimientos básicos de estadística, computación e inglés.

1.9. Arancel:

\$3500 (tres mil quinientos pesos argentinos). Los interesados pueden abonar la totalidad de la matrícula de inscripción hasta el 30 de octubre de 2020.

1.10. Fecha estimada de dictado (inicio-finalización). Modalidad a distancia.

Inicio: 2 de noviembre de 2020.

Finalización: 14 de diciembre de 2020.

Carga horaria total (a distancia): 75 hs distribuidas de la manera que se indica a continuación:

- **Modalidad a distancia (sincrónica obligatoria) total:** 30 hs. Las mismas estarán distribuidas en dos encuentros semanales de 2,5 horas/encuentro (5 horas semanales), durante 6 semanas. **Los horarios de estos encuentros serán: martes y jueves de 11.00 a 13.30 hs (horario Argentina).**

- **Modalidad a distancia con clases y/o actividades asincrónicas:** 25 hs. Las clases asincrónicas quedarán disponibles en la plataforma oportunamente indicada junto con la actividad práctica a realizar. La misma tendrá un plazo de entrega de 15 días.

- **Modalidad a distancia para realización del trabajo final (con tutorías a requerimiento):** 15 hs. El plazo de entrega de esta actividad se indicará durante el desarrollo del curso, al igual que la fecha estimada de devolución, calificación y entrega de certificados del curso.

2. Cuerpo Docente

2.1. Docente responsable

Dr. Luis Acuña Rello. Universidad de Valladolid. España.

Responsable del dictado general del curso y de tutorías a distancia; responsable del armado/corrección de la evaluación del curso. Carga horaria: 75 hs.

2.2. Docente co-responsable

Dra. Eleana Spavento. Facultad de Ciencias Agrarias y Forestales. UNLP. Argentina.

Responsable del desarrollo de los conceptos básicos introductorios sobre estadística-acercamientos al uso de R y participación en tutorías durante toda la duración del curso a distancia; participación en el armado evaluación/corrección de informes finales; responsable de la organización general del curso. Carga horaria: 75 hs.

3. Requisitos básicos para el dictado y participación

Disponibilidad de pc.

Acceso a internet durante todo el curso.