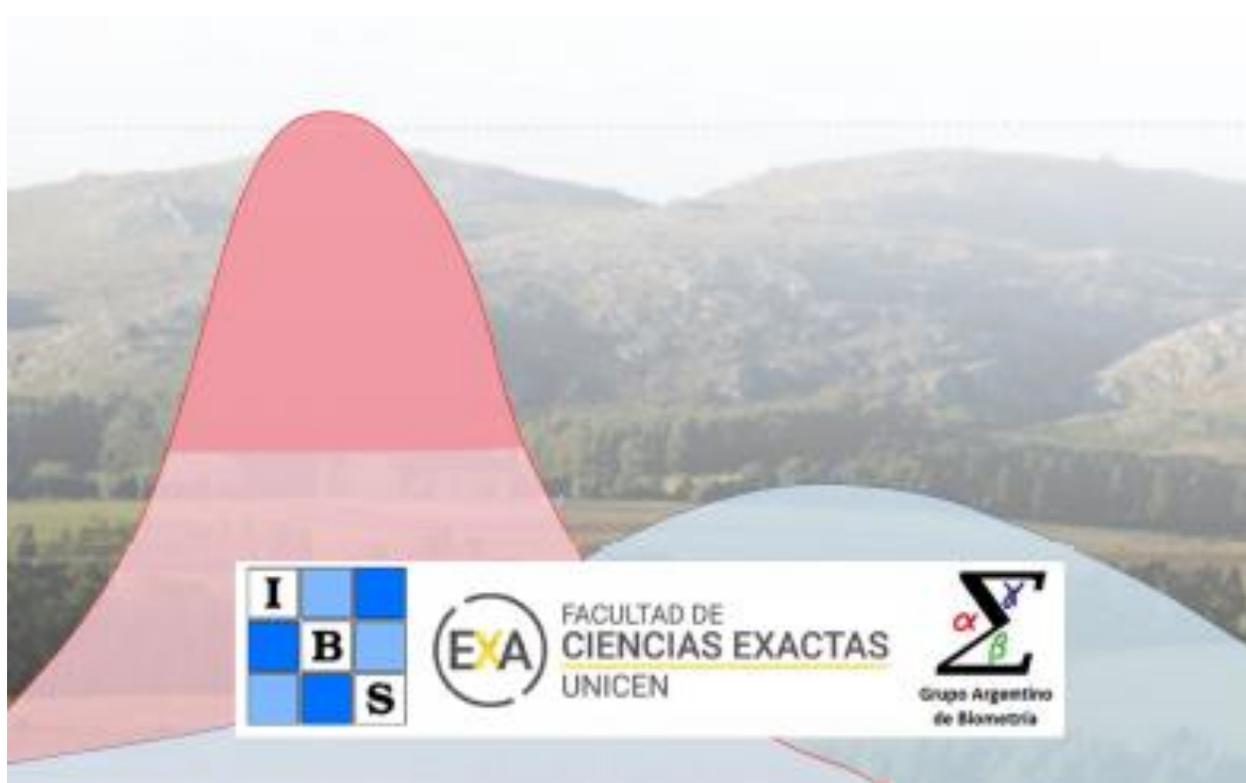


**XXV Reunión Científica  
del Grupo Argentino de Biometría**

10 al 12 de noviembre de 2020

Modalidad virtual

*Libro de Resúmenes*





**Grupo Argentino de Biometría**

Libro de Resúmenes. XXV Reunión Científica del Grupo Argentino de Biometría / compilado por Rosana Cepeda ... [et al.]. - 1a edición especial - Córdoba : Grupo Argentino de Biometría, 2021.

Libro digital, PDF

Archivo Digital: descarga

Edición para Grupo Argentino de Biometría

ISBN 978-987-23883-9-3

1. Bioestadísticas. 2. Estadísticas. 3. Geoestadística. I. Cepeda, Rosana, comp. II. Título.

CDD 519.5

**EDITOR:** Grupo Argentino de Biometría (GAB) – Región Argentina de la International Biometric Society.

**DISEÑO Y COMPAGINACIÓN:** Cepeda, Rosana; Marinelli, Claudia; Rodríguez, Edgardo; Rodríguez, Marcelo y Romero, María del Carmen.

ISBN 978-987-23883-9-3





# Contenido

Presentación .....	7
Conferencistas invitados .....	9
Comité Científico .....	12
Comité Evaluador .....	13
Comité Organizador .....	14
Comisión Directiva Región Argentina .....	15
Resúmenes .....	16
UNA PROPUESTA PEDAGÓGICA PARA LA ENSEÑANZA A DISTANCIA DE LA BIOMETRÍA BASADA EN LA TEORÍA DE LA INTERFAZ EDUCATIVA DE SCOLARI .....	17
ARCE OSVALDO E. A., RUIZ MARIA J., PÉREZ GONZALO A., GARCÍA MARÍA B., MANZUR JULIO .....	17
APLICACIONES DE MODELOS MIXTOS A DATOS DE HORMONA REPRODUCTIVA PROGESTERONA EN LAGARTOS SALVATOR MERIANAE .....	21
BANEGAS FLORENCIA C. , GARCIA VALDEZ VALERIA; CHAMUT SILVIA ; ARCE OSVALDO E. A.21	
AJUSTE DE CURVAS DE CRECIMIENTO EN FRUTOS DE NARANJA VALENCIA, MEDIANTE MODELOS NO LINEALES MIXTOS.....	25
BÓBEDA GRISELDA R. R., ZOILO CONSTANZA, GIMENEZ LAURA .....	25
VARIABILIDAD ESPACIO-TEMPORAL DE LA PRECIPITACIÓN EN EL LÍMITE OCCIDENTAL DE LA REGIÓN PAMPEANA. IMPLICANCIAS PARA LA AGRICULTURA .....	28
BOCA T.; CIPRIOTTI P.A.....	28
COMPARACIÓN DE MÉTODOS DE REGRESIÓN SIMBÓLICA PARA DATOS DE INTERVALO EN UN CONTEXTO BIOMÉTRICO .....	32
CARRIZO JORGELINA , MALLEA ADRIANA.....	32
UN ANÁLISIS AUTOMÁTICO DE DATOS DE SOBREVIDA MEDIANTE R .....	36
KURISS CLAUDIA CASTRO.....	36
ENSEÑANZA DE MATEMÁTICA Y BIOESTADÍSTICA EN MODALIDAD VIRTUAL .....	39
CASTRO MJ, NÚÑEZ M, PONTEVILLE CH.....	39
RALEO QUÍMICO, CAÍDA DE FRUTOS Y RETORNO FLORAL EN MANZANO RED CHIEF ..	42



CURETTI MARIELA, RAFFO DOLORES <sup>1</sup> , ROMA FERNANDO, RAGGIO FERNANDO, CASTRO ANDREA, CORIA LAUTARO, REEB PABLO .....	42
IDENTIFICACIÓN DE CLONES DE BANANA ( <i>MUSA SPP.</i> ) PARA LA CREACIÓN DE UN BANCO DE GERMOPLASMA.....	46
DEL MEDICO ANA PAULA, TENAGLIA GERARDO, VITELLESCHI MARÍA SUSANA, LAVALLE ANDREA, PRATTA GUILLERMO RAÚL .....	46
APLICACIÓN DE MODELOS MIXTOS NO LINEALES EN EL CRECIMIENTO DE MANZANAS CV. 'RED CHIEF' .....	50
DEL BRIO DOLORES, TASSILE VALENTÍN, REEB PABLO D. ....	50
APLICACIÓN DEL MODELO DE REGRESIÓN LOGÍSTICA PARA EL ESTUDIO DE LA DESERCIÓN ESTUDIANTIL EN LOS INGRESANTES 2019 DE LA FACULTAD DE CIENCIAS ECONÓMICAS, UNIVERSIDAD NACIONAL DE ENTRE RÍOS .....	54
D'IORIO STEFANÍA, AVILA OLGA B.....	54
IMPACTO DEL USO DE DISTINTOS ESTÁNDARES DE REFERENCIA PARA EL CÁLCULO DE LA PREVALENCIA DE RETRASO EN CRECIMIENTO .....	58
DRAMIS AGUSTÍN NICOLÁS, ALTSZYLER EDGAR, TURJANSKI PABLO GUILLERMO, CUETO GERARDO RUBÉN, FERNANDEZ MARÍA SOLEDAD, NUÑEZ PABLO, PEREZ ADRIANA ALICIA... ..	58
EXPLORACIÓN DE EFECTOS AMBIENTALES SOBRE EL TAMAÑO DE CAMADA EN TRES LÍNEAS DE CONEJOS .....	62
FERNÁNDEZ E.N., MARTINEZ R.D., ABBIATI N.N.....	62
MODELACIÓN DE C Y N PROVENIENTE DE BIOMASA DE RAÍCES EN FUNCIÓN DE LA PROFUNDIDAD DEL SUELO .....	66
GILI ADRIANA A., BELMONTE VALERIA, GARCÍA MARÍA V., FRASIER ILEANA, NOELLEMAYER ELKE J.....	66
APLICACIÓN DE REGRESIÓN LOGÍSTICA DE EFECTOS MIXTOS Y VALIDACIÓN CRUZADA PARA PREDECIR LA CAÍDA NATURAL DE FRUTOS EN PERAS DEL CULTIVAR "WILLIAMS" .....	70
GIMENEZ GUSTAVO, RUBIO NATALIA, BRAMARDI SERGIO .....	70
MODELANDO REGISTROS LONGITUDINALES DE CRECIMIENTO FÍSICO EN LA POBLACIÓN INFANTIL DE ARGENTINA: COMPARACIÓN DE MODELOS .....	74
GONZÁLEZ GARELLO TOMÁS, TURJANSKI PABLO, FERNANDEZ MARÍA SOLEDAD, NÚÑEZ PABLO, PEREZ ADRIANA ALICIA, CUETO GERARDO .....	74
DECODIFICANDO INFORMACIÓN EN REDES MULTICAPAS.....	78
HUAYLLA C., KUPERMAN M., NACIF M., COULIN C., GARIBALDI L. ....	78
ANÁLISIS FACTORIAL MÚLTIPLE PARA IDENTIFICACIÓN DE ASOCIACIONES ENTRE CARACTERES FENOTÍPICOS CUALITATIVOS Y MARCADORES MOLECULARES .....	82
LAC PRUGENT V.; COSTA TÁRTARA S.; CURTI R.; BRAMARDI S.....	82



EL EFECTO DE LA DISTANCIA MARÍTIMA EN EL COMERCIO DE PRODUCTOS PESQUEROS: ANÁLISIS GRAVITACIONAL.....	86
LACAZE MARÍA VICTORIA .....	86
ESTRATEGIA DE MUESTREO CON INFORMACIÓN ESPACIAL EN PRONÓSTICOS DE PRODUCCIÓN DE VID .....	90
LLERA J., MARTINENGO N., MORELLI CL., MARTÍNEZ E., MAURE E., RODRÍGUEZ PLAZA L., CARRILLO N., CORVALÁN, A.; GALIOTTI, H.; NAZRALA, J.; SOSA, M. ....	90
MEDICINA BASADA EN LA EVIDENCIA: PRUEBAS DIAGNÓSTICAS Y SU APLICACIÓN EN ENSAYOS DE EVALUACIÓN DE DESEMPEÑO DE UN TEST RÁPIDO DE DETECCIÓN DE COVID 19.....	94
LOPEZ VANESA, DELFINO VANINA, FILIPPINI O. SUSANA.....	94
PATRONES ESPACIALES EN INDICADORES DE MALNUTRICIÓN INFANTIL DE LA REPÚBLICA ARGENTINA .....	98
MELCONIAN M.A, FERNANDEZ M.S, TURJANSKI P., ALTSZYLER E., LIBMAN M, PEREZ A., NUÑEZ P., CUETO G.R.....	98
ANÁLISIS EXPLORATORIO DE LAS CONDUCTAS ALIMENTARIAS: FACTORES QUE INCIDEN EN LAS ELECCIONES DE LOS ESTUDIANTES UNIVERSITARIOS.....	101
MENÉNDEZ LUCIANO, BERGES MIRIAM, LUPÍN BEATRIZ .....	101
CONTAMINACIÓN VEHICULAR. MODELADO GEOESTADÍSTICO CON VARIOGRAMAS DIFUSOS .....	105
MOLINARI DANIELA A., CHAPARRO MAURO A.E., CHAPARRO MARCOS A.E. ....	105
RENDIMIENTO ACADÉMICO EN ÓPTICA GEOMÉTRICA .....	109
MONTENEGRO BRUSOTTI JUDITH, PONTEVILLE CHRISTIANE, NÚÑEZ MYRIAM.....	109
ALTERNATIVAS PARA MEDIR LA VARIABILIDAD ESPACIAL DE LA MATERIA ORGÁNICA DEL SUELO A PARTIR DE SENSORES VERSUS MUESTREOS DE LABORATORIO .....	112
MUÑOZ H. M., ZABALA S. M., VARGAS S. L., FILIPPINI O. S., VÁZQUEZ J. M.....	112
INVESTIGACIÓN DEL HÁBITO TABÁQUICO EN PACIENTES ONCOLÓGICOS, EN ENFERMEDADES NEOPLÁSICAS CON Y SIN RELACIÓN CON EL CONSUMO DE CIGARRILLOS .....	116
NUÑEZ M., BONADIES E., VERZURA M.A., MAUR PEROTTI J.A., PEREZ MESA P., SPECTERMAN S., LASTIRI J.M.....	116
UNA CHECKLIST PARA MAXIMIZAR LA REPRODUCIBILIDAD EN ESTUDIOS DE DIVERSIDAD FUNCIONAL.....	120
PALACIO FACUNDO X.....	120
EVALUACIÓN DE ALTERNATIVAS AL USO DE LOS ANTIBIÓTICOS EN ENSAYOS AVÍCOLAS UTILIZANDO MODELOS LINEALES MIXTOS .....	124
PALACIOS LUCIANO; FILIPPINI OLGA SUSANA; DE FRANCESCHI MAURICIO .....	124



---

ANÁLISIS ECONÓMETRICO DE LA VALORACIÓN DE ACEITE DE OLIVA VIRGEN EXTRA .....	127
ROLDÁN CAMILA, LUPÍN BEATRIZ, TEDESCO LORENA.....	127
MODELO BINOMIAL NEGATIVA PARA AJUSTAR LA PRESENCIA DE CHIMANGOS EN EL AEROPUERTO INTERNACIONAL DE LA CIUDAD DE NEUQUÉN .....	131
RUBIO NATALIA, ANGELICCHIO JUAN, GIMENEZ GUSTAVO N.....	131
ESTUDIO DEL RENDIMIENTO ACADÉMICO DE ESTUDIANTES DE CONTADOR PÚBLICO NACIONAL.....	135
SABINO GUILLERMO, ARENAS LUIS.....	135
COMPETENCIAS DOCENTES EN DOCTORES EN CIENCIAS BIOLÓGICAS, ANÁLISIS DESDE SUS PERCEPCIONES .....	139
SIMONIELLO M. FERNANDA, ODETTI HECTOR, TEMPORETTI FELIX, AVILA OLGA, CONTINI LILIANA .....	139
ANÁLISIS DE LAS CARACTERÍSTICAS HÍDRICAS Y CLIMÁTICAS EN LA CUENCA DEL RÍO COLLÓN CURÁ, MEDIANTE MODELOS LINEALES GENERALIZADOS.....	143
ULLOA DIEGO, HAIQUE ANA KARINA, CARRASCO GABRIELA, RABY GABRIEL .....	143
ELIMINACIÓN DE NUTRIENTES Y MATERIA ORGÁNICA EN AGUA RESIDUAL URBANA POR HUMEDALES CONSTRUIDOS .....	146
VAIRA STELLA, CARRIO JOSEFINA, PIZARRO ANA, ZERBATTO MARIEL, MODINI LAURA .....	146
BÚSQUEDA DE ESTRUCTURA GENÉTICA: COMPARACIÓN DE ALGORITMOS .....	150
VIDELA M EUGENIA, BRUNO CECILIA .....	150
ANÁLISIS CANÓNICO DE CORRESPONDENCIAS Y ANÁLISIS DE COINERCIAS: UNA APLICACIÓN EN EL CAMPO DE LA ECOLOGÍA.....	154
ZULIANI ROXANNA, LAFFITTE LORENA, MUÑIZ SAAVEDRA JULIETA, LAVALLE ANDREA.....	154

## Presentación

La International Biometric Society es una sociedad internacional que promueve el desarrollo y la aplicación de la teoría y los métodos estadísticos y matemáticos en las biociencias, incluidas la agricultura, la ciencia biomédica y la salud pública, la ecología, las ciencias ambientales, la silvicultura y disciplinas afines.

La Región Argentina de la International Biometric Society (IBS) llevó a cabo la XXV Reunión Científica del Grupo Argentino de Biometría (GAB) los días 10 a 12 de Noviembre de 2020, organizada por el Grupo Argentino de Biometría y la Facultad de Ciencias Exactas de la Universidad Nacional del Centro de la Provincia de Buenos Aires (UNICEN, Tandil). En el marco de la emergencia por Covid 19, esta reunión se realizó, por primera vez, en modalidad virtual.

El encuentro tuvo como objetivo proveer un ámbito propicio para el intercambio y la discusión de ideas motivadoras e inspiradoras para la comunidad de profesionales especializados en Biometría, difundir los resultados de su trabajo, promover y fortalecer la generación de vínculos entre instituciones y reforzar la formación de recursos humanos. Las actividades programadas fueron: conferencias de reconocidos biometristas nacionales e internacionales, mini-cursos, exposición de trabajos científicos, mesa de análisis de datos y la Asamblea Anual Ordinaria del GAB.

### Minicursos

- “Cómo diseñar un informe de manera rápida y sencilla con R Markdown”, a cargo de la Dra. Teresa Boca.
- “Modelos Estadísticos en Mejoramiento Genético”, a cargo de la Dra. Cecilia Bruno.

### Conferencias y charlas

- Conferencia inaugural: “Modelos estadísticos y epidemiológicos aportando al conocimiento de la pandemia COVID-19 en la Región”. Dra. Gloria Icaza Noguera (Universidad de Talca, Chile) y Dra. María del Pilar Díaz (Universidad Nacional de Córdoba, Argentina).
- Conferencia: “Reglas para la selección del conjunto de variables de ajuste óptimo en modelos gráficos causales”. Dra. Andrea Rotnitzky (Universidad Torcuato Di Tella, Argentina).
- Conferencia: “Meta-análisis Bayesiano”. PhD. Pablo Verde (Universidad de Duesseldorf, Alemania).



- Conferencia: “Desarrollo de un pipeline para selección genómica utilizando grandes matrices de marcadores (3K genotipos y 14 millones de marcadores) en garbanzo”. PhD. Diego Jarquin (Universidad de Nebraska –Lincoln, USA).
- Charla: “Modelos decisionales inteligentes para análisis en Big Data”. Lic. Jorge Sagula (Universidad Nacional de Luján).

## Conferencistas invitados

**Dra. Gloria Icaza Noguera (Universidad de Talca, Chile) y Dra. María del Pilar Díaz (Universidad Nacional de Córdoba, Argentina)**

Conferencia: “Modelos estadísticos y epidemiológicos aportando al conocimiento de la pandemia COVID-19 en la Región”

En esta charla se abordará el estado de situación en relación al covid, señalando el contexto de Chile, se muestran las estadísticas generales, modelos aplicados desde organismos del estado y análisis de la situación actual. Se muestra además el diseño e implementación del estudio de seroprevalencia en población de base. Finalmente se mencionan los principales desafíos para el estudio de mortalidad. Por otra parte, en relación a la situación de Argentina, menciona situación actual y el contexto de Córdoba se presentan estadísticas generales, modelos de incidencia así como modelos para la estimación del número reproductivo básico ( $R_0$ ) y Alternativas adaptadas al caso de Córdoba.

**Dra. Andrea Rotnitzky (Universidad Torcuato Di Tella, Argentina)**

Conferencia: “Reglas para la selección del conjunto de variables de ajuste óptimo en modelos gráficos causales”

Un concepto causal es una relación que no puede ser definida a partir solamente de la distribución conjunta de las variables observadas. Así las relaciones causales requieren introducir variables que ayuden a conceptualizar el problema de interés. En esta charla se considera la selección de posibles variables de confusión en la etapa del diseño de un estudio observacional planificado. Dado un modelo causal gráfico no paramétrico tentativo, el objetivo será seleccionar el conjunto de covariables que sean suficientes para controlar la confusión bajo el modelo y que produzcan un estimador no paramétrico del contraste causal de interés con la menor varianza. Desarrollaremos reglas gráficas para determinar el conjunto de ajuste óptimo, basados en la teoría de Grafos.

**PhD. Diego Jarquín (Universidad de Nebraska –Lincoln, USA)**

Conferencia: “Desarrollo de un pipeline para selección genómica utilizando grandes matrices de marcadores (3K genotipos y 14 millones de marcadores) en garbanzo”

La selección genómica es una metodología que vino a revolucionar la forma convencional de mejorar cultivos con el uso de la información molecular. Los métodos tradicionales de mejoramiento genético consideran el uso de información fenotípica lo cual implica la observación de genotipos en campo. En muchas ocasiones es complicado, si no es que imposible el observar todos los genotipos (candidatos) en campo para de ahí seleccionar el mejor o los mejores para los siguientes ciclos de mejora. Las restricciones impuestas por el costo, la disponibilidad de tierra, agua y semilla complican la observación de todos los materiales genéticos en campo. Sin embargo, siempre habrá la oportunidad para observar una proporción de estos. El uso de información molecular puede ser útil para predecir el desempeño esperado de los genotipos no observados. De esta forma, los mejoradores usarán los valores predichos como reemplazo de los valores observados al momento de realizar la selección de genotipos. En este tipo de problemas de predicción, por lo regular el número de marcadores moleculares (algunos miles o decenas de miles,  $p$ ) sobrepasa el número de genotipos (algunos cientos,  $n$ ). Varios modelos han sido desarrollados para sortear las complicaciones de invertir matrices que no son de rango completo (cuando  $p \gg n$ ), las cuales se emplean en la obtención de los valores predichos. El esquema desarrollado puede aplicarse fácilmente a problemas de otros campos de estudio que envuelvan el uso de matrices de grandes dimensiones.

**PhD. Pablo Verde (Universidad de Duesseldorf, Alemania)**

Conferencia: “Meta-análisis Bayesiano”

Los investigadores de salud pública pueden tener que decidir si realizar un metanálisis que incluya solo ensayos clínicos aleatorios (ECA) de alta calidad o si incluir una combinación de todas las pruebas disponibles, es decir, ECA de calidad variable y estudios observacionales (SG). El principal obstáculo al combinar pruebas dispares en un metanálisis es que no solo combinamos resultados de interés, sino que también combinamos múltiples sesgos. Por lo tanto, los métodos de metanálisis comúnmente aplicados pueden llevar a conclusiones engañosas. En este trabajo, presentamos un nuevo modelo jerárquico bayesiano, llamado modelo de metanálisis con corrección de sesgo (BC), para combinar diferentes tipos de estudios en el metanálisis. Este modelo se basa en una mezcla de dos distribuciones de efectos aleatorios, donde el primer componente corresponde al modelo de interés y el segundo componente a la estructura

de sesgo oculto. De esta forma, el modelo de interés resultante se ajusta por el sesgo de validez interna de los estudios incluidos en una revisión sistemática.

### **Lic. Jorge Sagula (Universidad Nacional de Luján)**

Conferencia: Modelos decisionales inteligentes para análisis en Big Data

*Big Data* es un conglomerado de proporciones que engloba un conjunto de modelos, metodologías, técnicas y tecnologías especializadas en aras de procesar voluminosos conjuntos de datos, y consecuentemente, debido al gran volumen y a la complejidad misma no es suficiente emplear exclusivamente el herramental disponible de administración de bases de datos. El análisis de datos corresponde al proceso de estructuración de *Big Data*, y así, la existencia de diferentes patrones y relaciones permiten una mejor caracterización de los datos, de tal modo el análisis de datos en sí mismo reviste alta relevancia en la Tecnología de la Información.

Los modelos de Inteligencia Artificial contribuyen con vastas teorías, técnicas y metodologías a las Ciencias de la Computación, hoy en particular a *Big Data*, tales como: Reglas de Asociación, Algoritmos Genéticos, Análisis de Regresión, Análisis de Clustering, Machine Learning, Modelos de Análisis de Redes Sociales, Redes Neuronales, Data Mining, Deep Learning, Series Temporales, etc. Mejorar la capacidad de “*Toma de Decisiones*” corporativas, requiere de un herramental de calidad al efecto, que no sólo permita modelar adecuadamente, sino fundamentalmente, predecir comportamientos futuros, y es aquí donde el Aprendizaje Estadístico toma una preponderante posición.



## Comité Científico

Balzarini, Mónica

Bramardi, Sergio

Casanoves, Fernando

Cendoya, Gabriela

Di Rienzo, Julio

Filippini, Susana

Macchiavelli, Raúl

Reeb, Pablo

Winzer, Nélica



## Comité Evaluador

Boca, Teresa

Bruno, Cecilia

Giménez, Laura

Lavalle, Andrea

Monterubbianesi, Gloria

Pérez, Adriana

Saravia, José

Sührling, Silvia

Zuliani, Paola



## Comité Organizador

Cepeda, Rosana

Marinelli, Claudia

Rodríguez, Edgardo

Rodríguez, Marcelo

Romero, María del Carmen



## **Comisión Directiva Región Argentina**

**Presidente:** Pablo Daniel Reeb

**Vicepresidente:** Silvia Sühling

**Secretario:** María del Carmen Romero

**Tesorero:** Teresa Boca

**Vocal Titular:** Pablo Demetrio

**Vocal Titular:** Adriana Perez

**Vocal Suplente:** Liliana Contini

**Vocal Suplente:** Mariel Vanin



# Resúmenes

## **UNA PROPUESTA PEDAGÓGICA PARA LA ENSEÑANZA A DISTANCIA DE LA BIOMETRÍA BASADA EN LA TEORÍA DE LA INTERFAZ EDUCATIVA DE SCOLARI**

ARCE OSVALDO E. A., RUIZ MARIA J., PÉREZ GONZALO A., GARCÍA MARÍA B., MANZUR JULIO<sup>1</sup>

<sup>1</sup>*Cátedra Biometría y Técnica Experimental, Facultad de Agronomía y Zootecnia, Universidad Nacional de Tucumán  
oarce@faz.unt.edu.ar*

### **RESUMEN**

En este trabajo se presenta una propuesta pedagógica para la enseñanza de la asignatura Biometría y Técnica Experimental de la Facultad de Agronomía y Zootecnia (FAZ) de la Universidad Nacional de Tucumán (UNT), en el contexto del aislamiento social obligatorio, que impidió el dictado de la asignatura en la modalidad presencial tradicional y que obligó a repensar la práctica docente a través de la modalidad a distancia mediada por la tecnología. Se encuadra esta propuesta dentro de la teoría de la interfaz educativa de Scolari. Esta teoría se consideró apropiada ya que permite establecer interrelaciones multidireccionales entre los actores humanos y tecnológicos de la red garantizando procesos de significación y coevolución. Se presentan además los resultados de una encuesta hecha a los alumnos que participaron de la propuesta en el cursado 2020 de la asignatura como validación de la propuesta.

**Palabras clave:** *didáctica estadística, red aprendizaje, enseñanza virtual.*

### **INTRODUCCIÓN**

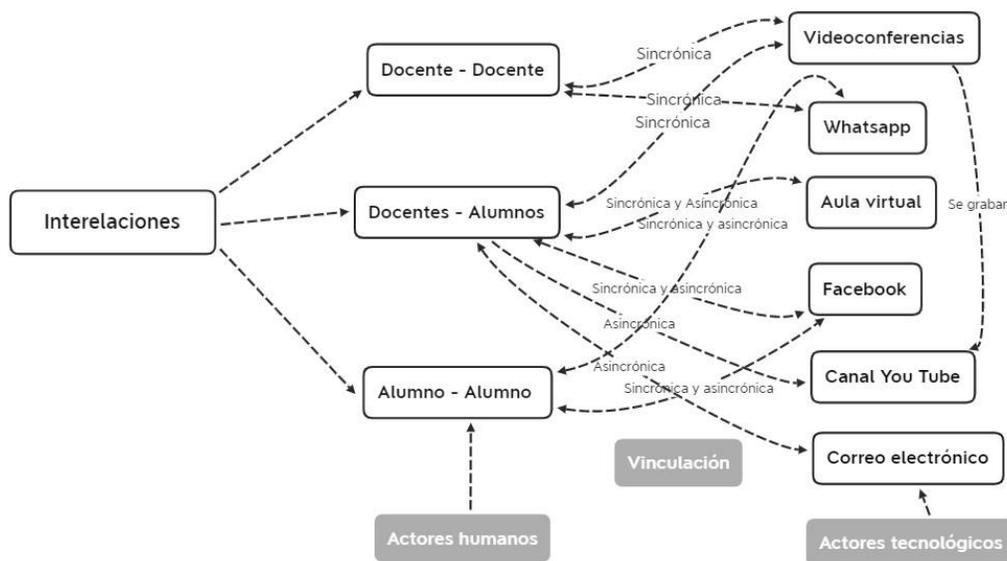
La repentina aparición de la pandemia de COVID19 obligó a los docentes a tomar decisiones rápidas sobre cómo dictar, qué metodologías usar, qué programas y plataformas, garantizando la calidad del proceso de enseñanza-aprendizaje en un contexto de no presencialidad. Según el mapa de Bates y Poole (2013) se tuvo que pasar en forma abrupta de un extremo al otro (presencialidad total a virtualidad total). En general, la tradición presencial no consideró herramientas tecnológicas aun conociendo el hecho de que los estudiantes las manejan en su vida social y a pesar de que están disponibles desde hace mucho tiempo (Gómez, 2020). La base de la propuesta se basa en estos dos conceptos: “no es la modalidad (presencial o virtual) lo que garantiza mejores clases y procesos de enseñanza basados en el desarrollo del pensamiento crítico; es la propuesta político-pedagógica del docente la que lo define” (Juarros y Levy, 2020) y “Saber enseñar no es transferir conocimiento, sino crear condiciones para su propia producción o construcción” (Freire, 1996). La tecnología es un medio invaluable para complementar el proceso de enseñanza aprendizaje, pero es la creatividad en cómo usarla lo que la hace potente para la enseñanza (Dussel, 2018). Aprender con tecnologías ofrece oportunidades cognitivas diferentes a las que se generan al aprender sin ellas. No es posible reproducir la modalidad presencial, sino que todo el proceso debe repensarse, generando “nuevas formas de planificar, interpretar y comprender el rol docente y la clase misma” (Martín,

2020). Scolari (2019) propone una resignificación del concepto de interfaz como interrelación individuo – computadora, a uno más general en el cual “una interfaz es una red de actores humanos y tecnológicos que interactúan y mantienen diferentes tipos de relaciones entre sí”. En este contexto una interfaz educativa, con interacciones multidireccionales entre los actores, mediada por la tecnología, es importante en el diseño de una propuesta pedagógica que no solo transmita contenidos, sino que además permita la comunicación permanente docente-alumno por diversos medios y el seguimiento cercano del proceso de adquisición de conocimientos por parte de los estudiantes.

El objetivo del trabajo fue evaluar la efectividad de la aplicación de la interfaz educativa de Scolari en el dictado a distancia de la asignatura Biometría y Técnica Experimental en la Facultad de Agronomía y Zootecnia de la Universidad Nacional de Tucumán.

## MATERIALES Y MÉTODOS

Se diseñó una propuesta pedagógica basada en el concepto de interfaz educativa de Scolari (2019). El esquema resume la vinculación entre actores, las modalidades e interrelaciones.



La propuesta se aplicó durante todo el cursado de la asignatura Biometría y Técnica Experimental de la FAZ de la UNT, en el primer cuatrimestre de 2020, desde marzo hasta junio. Se publicó semanalmente en el tablón de novedades de *Classroom* el cronograma de actividades de la semana siguiente con especificación de días y horarios de clases teórico-prácticas, prácticas y de consulta *on line* a través de videoconferencias. Se subía junto con el cronograma todo el material de lectura en formato pdf. El material es de autoaprendizaje con teoría explicada en detalle, ejercicios resueltos y ejercicios sin resolver para que los alumno/as lo trabajen solos/as usando los conceptos vertidos. El alumno/a debía leer este material. Las clases teórico-prácticas se dieron siempre en los mismo días y horarios de manera de generar una regularidad del estudiante en su dedicación a la materia. Fue una actividad sincrónica en la que se trabajaba el material previamente subido la semana anterior. Se explicaba el material y se abrían espacios de discusión con los alumnos y el profesor/a sobre dudas durante el estudio. Estas sesiones se grabaron y se subieron al canal de YouTube Biometria FAZ para alumnos

con problemas de acceso o conectividad y para que los/as alumnos/as que estuvieron en la actividad pudieran ver y escuchar nuevamente la clase las veces que desean. Consideramos que el alumno también necesitaba clases más personalizadas en las cuales pudiera evacuar sus dudas personales y sus dificultades puntuales al enfrentar el proceso de aprendizaje. Para ello se programaron clases de consulta por videoconferencia en donde se contestaron las preguntas de los alumnos usando el material entregado, en conjunto con los profesores. Se brindaron diferentes horarios y días para dar la oportunidad de participación a todos/as. La semana posterior a las clases se realizó un control de lectura para cada tema, con devolución inmediata de resultados para que el/la alumno/a pudiera seguir su proceso de aprendizaje. Se hizo en formularios de Google en formato cuestionario de autocorrección. También se tomaron dos evaluaciones cuantitativas integradoras para obtener la promoción o la regularidad. con preguntas de selección múltiple. Al finalizar el cursado se hizo una encuesta para evaluar la efectividad de la propuesta desde la perspectiva del estudiante.

## **RESULTADOS Y DISCUSIÓN**

Cursaron 152 alumnos/as. Participaron activamente a través de todos los medios tecnológicos disponibles puestos a su disposición. Principalmente utilizaron las videoconferencias para clases presenciales y consultas personalizadas (en promedio 30 alumnos por clase). Los videos de las clases en el canal de YouTube tuvieron en promedio 300 reproducciones lo cual indicó un aprovechamiento importante de las clases asincrónicas. El uso del correo electrónico como fuente de consulta fue permanente. Se atendieron 460 consultas por ese medio. En los controles de lectura (evaluación cualitativa) participaron la totalidad de los estudiantes. Facebook y WhatsApp fueron poco elegidos, sólo para consultas muy puntuales. No obstante, sirvieron para comunicaciones masivas que debían llegar rápidamente a todo el grupo, como cambio de horario de una clase por ej. Las evaluaciones cuantitativas (parciales) arrojaron los siguientes resultados: 46 promocionados, 63 regulares y 43 libres.

En la encuesta el 92 % manifestó que la propuesta le resultó útil; 93 % consideró que los materiales de autoaprendizaje puestos a su disposición fueron claros y que las clases grabadas les fueron útiles; el 67 % siguió el proceso por celular; el 86 % consideró que nuestra propuesta merecía una calificación de entre 8 y 10.

Algunas frases de alumnos en esta encuesta: “agradecerles por la buena onda y la predisposición de siempre y decirles que son la única materia que se supo organizar y pensó en nosotros los alumnos. De todo corazón muchísimas gracias, se nota en cada clase la dedicación y el amor por lo que hacen”; “con toda la predisposición que tiene la cátedra para que el cursado sea más liviano, realmente es más fácil llevar la materia al día, temas que no entendía, que me costaban entender, ahora es más fácil. Gracias a todos por su tiempo”; “muchas gracias por el gran esfuerzo que realiza la cátedra para con nosotros, ha sabido usar bien la implementación de la tecnología en este tiempo de emergencia”; “me parece que está muy bien la forma de cursada. Al no entender algo puedo ver nuevamente los videos, además los profesores están siempre a disposición”; “el único aspecto negativo que noto es que a veces hay problemas de conexión, pero entiendo que son temas que no son culpa de los profesores. Sin embargo, las clases luego las suben a YouTube, lo cual es genial”. Y muchas más de ese tenor.

Consideramos que la interfaz educativa de Scolari es un marco apropiado para planificar el dictado de Biometría en forma no presencial.

## **REFERENCIAS**



- [1] Bates, A. y Poole, G. (2003). *Effective Teaching with Technology in Higher Education* Jossey-Bass/John Wiley, EEUU.
- [2] Dussel, Inés (2018). “¿Nuevas formas de enseñar y aprender?” *Perfiles Educativos*, XL:143-178.
- [3] Freire, Paulo (1996). *Pedagogía de la autonomía. Saberes necesarios para la práctica educativa*. Siglo XX, Argentina.
- [4] Gómez, P. A. (2020). *Cultura de la conectividad y subjetividades actuales en la relación estudiantes - docentes. Pedagogía crítica y didáctica en la enseñanza digital*. Ministerio de Educación de la Nación, Argentina.
- [5] Juarros, M.F. y Levy, E. (2020). *La práctica docente en la educación a distancia. La relación pedagógica mediada por tecnologías. Pedagogía crítica y didáctica en la enseñanza digital*. Ministerio de Educación de la Nación, Argentina.
- [6] Martín, M.M. (2020). *Perspectivas pedagógico-didácticas en la enseñanza universitaria en entornos virtuales. Pedagogía crítica y didáctica en la enseñanza digital*. Ministerio de Educación de la Nación, Argentina.
- [7] Scolari, C. A. (2019). “¿Cómo analizar una interfaz?”. *Method*, Cataluña. Versión obtenida el 02/09/2020. [www.researchgate.net/publication/330651740\\_Como\\_analizar\\_una\\_interfaz](http://www.researchgate.net/publication/330651740_Como_analizar_una_interfaz)

## APLICACIONES DE MODELOS MIXTOS A DATOS DE HORMONA REPRODUCTIVA PROGESTERONA EN LAGARTOS *SALVATOR MERIANAE*

BANEGAS FLORENCIA C.<sup>1</sup>, GARCIA VALDEZ VALERIA<sup>2</sup>; CHAMUT SILVIA<sup>2</sup>; ARCE OSVALDO E. A.<sup>1</sup>

<sup>1</sup>Cátedra Biometría y Técnica Experimental, Facultad de Agronomía y Zootecnia, UNT

<sup>2</sup>Cátedra Biología del Desarrollo, Facultad de Agronomía y Zootecnia, UNT  
flor\_caro\_8@hotmail.com

### RESUMEN

Numerosos experimentos agropecuarios están basados en experimentos factoriales en los cuales una combinación de factores se asigna a cada unidad experimental y luego se toman medidas repetidas en el tiempo dentro de cada una de ellas. Este trabajo propone la modelación estadística de los datos de la hormona progesterona en lagartos *Salvator merianae* criados en cautiverio. Como criterios para la selección del modelo se tuvo en cuenta el valor del criterio de información de Akaike y la variabilidad residual. El modelo de mejor ajuste correspondió al que tenía el año como efecto aleatorio y una función exponencial para el modelado de la varianza.

**Palabras clave:** *modelos lineales, datos autocorrelacionados, medidas repetidas en el tiempo.*

### INTRODUCCIÓN

Numerosos trabajos agropecuarios están basados en experimentos factoriales en los cuales una combinación de factores se asigna a cada unidad experimental y luego se toman medidas repetidas en el tiempo dentro de cada una de ellas. Este tipo de datos se llaman longitudinales. Los modelos de población promedio se refieren a la expectación de la variable respuesta como el interés principal de análisis. Entre las propuestas sobre la modelación de estos datos se encuentra la de modelos mixtos lineales (Pinheiro y Bates, 2000; West *et al.*, 2015; Wu, 2010). Las variaciones estacionales en el comportamiento del lagarto *Salvator merianae* son impulsadas por cambios en las hormonas que afectan los distintos momentos del ciclo anual: salida de hibernación, cortejo, vitelogénesis, oviposición, incubación y quiescencia reproductiva. Así también existe un efecto hormonal en el ovario que determina el ciclo folicular (ciclo folicular ovulatorio y ciclos foliculares anovulatorios de cursos corto y largo). La hormona progesterona es una de las involucradas en regular dichos ciclos (García Valdez *et al.*, 2016; Jones, 2011). Los datos provenientes de este estudio están caracterizados por tener falta de independencia entre las observaciones debida a dos causas: las medidas repetidas en el tiempo para cada animal dentro de cada año y a través de los años ya que los animales se mantienen en criadero por varios años. Por otra parte, se presenta heterocedasticidad en las mediciones. Este trabajo propone la modelación estadística, mediante modelos lineales mixtos, de datos de la hormona progesterona en lagartos *S. merianae*, criados en cautiverio, a lo largo del ciclo anual/estacional.

## MATERIALES Y MÉTODOS

Para el modelado estadístico se utilizó el programa R (R Core Team, 2020) y sus paquetes *nlme*, *DAAG* y *emmeans*. Se evaluaron los efectos fijos: Momento, Ciclo y Momento x Ciclo. Se probaron distintas estructuras para la parte aleatoria: año, animal, animal anidado dentro del año y animal cruzado con el año. Se utilizaron como criterios para la selección de modelos: criterio de información de Akaike (AIC) y la variabilidad residual. Se consideró como mejor modelo a aquél que tuviera los valores de AIC y de error más bajos. Además, se verificaron los supuestos para los términos de error (normalidad, homogeneidad de varianza e independencia). Para la heterocedasticidad se ajustaron diferentes estructuras de varianza que fueron: heterocedasticidad por momentos del ciclo, función de potencia de varianza y función exponencial. Con el paquete *emmeans* de R se compararon las medias marginales estimadas a partir del modelo seleccionado.

## RESULTADOS Y DISCUSIÓN

Los modelos que mejor ajustaron fueron los que presentaron como efectos aleatorios año y animal anidado dentro de año con función exponencial para la modelar la varianza residual (filas sombreadas en la tabla 1). Si bien éstos no son los de menor AIC sí presentaron una reducción importante en la variabilidad residual como se observa en la tabla 1. La prueba de razón de verosimilitud para estos dos modelos no mostró diferencias significativas ( $p$ -valor = 0.2424). Por el principio de parsimonia se estableció que el modelo más simple es el más apropiado al tener menos parámetros estimados. En este caso corresponde al modelo con efecto aleatorio del año.

Tabla 1. Resumen de modelos ajustados

Efecto aleatorio	Estructura de varianza	AIC	Sigma estimado
Año	Constante	277.9	0.5311
Año	Heterocedasticidad por Momento	232.8	0.4057
Año	Función de potencia	211.4	0.5413
Año	Función Exponencial	221.1	0.1352
Animal	Constante	302.9	0.5689
Animal	Heterocedasticidad por Momento	254.4	0.4639
Animal	Función de potencia	245.4	0.6210
Animal	Función Exponencial	No converge	No converge
Animal dentro de año	Constante	278.5	0.5087
Animal dentro de año	Heterocedasticidad por Momento	229.8	0.4135
Animal dentro de año	Función de potencia	207.2	0.5467
Animal dentro de año	Función Exponencial	221.8	0.1277
Animal Cruzado con año	Constante	293.5	0.5140
Animal Cruzado con año	Heterocedasticidad por Momento	232.8	0.4057
Animal Cruzado con año	Función de potencia	225.2	0.6025
Animal Cruzado con año	Función Exponencial	No converge	No converge

La tabla 2 muestra que la prueba de Wald para el efecto principal Momento es el único significativo.

Tabla 2. Prueba de Wald para los efectos fijos

	gl Numerador	gl Denominador	F	p-valor
(Intercept)	1	118	11.584	0.0009
Momento	6	118	10.623	<.0001
Ciclo	2	118	1.412	0.2477
Momento: Ciclo	12	118	1.038	0.4186

La tabla 3 muestra los efectos estimados y su significación para los distintos momentos del ciclo. Se observa que los efectos más importantes son: entrada hibernación, hibernación y salida hibernación.

Esta metodología de análisis permitió modelar la falta de independencia y la heterocedasticidad de los términos de error, siendo muy útil para el análisis de datos zootécnicos que normalmente presentan estas características. Podemos concluir también que no debería usarse solamente el AIC como criterio de mejor ajuste de un modelo, sino que además se deben tener en cuenta la estimación de la variabilidad aleatoria y el cumplimiento de los supuestos para los términos de error.

Tabla 3. Efectos estimados para Momento

Contraste	Estimación	E.E.	gl	t.ratio	p-valor
Efecto Cortejo	0.5270	0.1842	118	2.860	0.0345
Efecto Entrada Hibernación	-0.4713	0.0718	118	-6.561	<.0001
Efecto Hibernación	-0.3892	0.0770	118	-5.054	<.0001
Efecto Incubación	0.1900	0.1217	118	1.562	0.5946
Efecto Oviposición	0.5095	0.1933	118	2.636	0.0648
Efecto Quiescencia	-0.0571	0.0986	118	-0.579	0.9970
Efecto Salida Hibernación	-0.3089	0.0680	118	-4.546	0.0001

## REFERENCIAS

- [1] García Valdez, V.; Chamut, S.; Jahn, G.; Arce, O.; Manes, M. (2016). "Plasmatic estradiol and progesterone variations during the reproductive cycle of captive female lizard *Tupinambis Rufescens*". *Herpetological Conservation and Biology*. Vol.11: 519 - 526.
- [2] Jones, S.M. (2011). "Hormonal regulation of ovarian functions in reptiles". Publicado en Norris D.; Lopez K. (2011), *Hormones and Reproduction of Vertebrates*. Vol 3. Elsevier, USA, 89-115
- [3] Pinheiro J.C., Bates D.M. (2000). *Mixed-Effects Models in S and S-PLUS*. Springer-Verlag, USA.
- [4] R Core Team (2020). "R: A language and environment for statistical computing. R Foundation for Statistical Computing". Vienna, Austria. Versión obtenida el 01/08/2020. <http://www.r-project.org/>.



- [5] West B., Welch K., Galecki A. (2015). *Linear mixed models. A practical guide using statistical software*. 2nd edition. Chapman & Hall/CRC, USA.
- Wu L. (2010). *Mixed effects models for complex data*. Chapman & Hall/CRC, USA.

## AJUSTE DE CURVAS DE CRECIMIENTO EN FRUTOS DE NARANJA VALENCIA, MEDIANTE MODELOS NO LINEALES MIXTOS

BÓBEDA GRISELDA R. R.<sup>1</sup>, ZOILO CONSTANZA<sup>1</sup>, GIMENEZ LAURA<sup>1</sup>

<sup>1</sup>Cátedra Cálculo Estadístico y Biometría 1, Departamento Matemáticas, Facultad de Ciencias Agrarias-Universidad Nacional del Nordeste  
griseldabobeda@gmail.com

### RESUMEN

El crecimiento de frutos de naranjo dulce responde a una curva sigmoidea en forma de S, no lineal en sus parámetros. Para su ajuste se utilizan modelos de regresión no lineal. El estudio consistió en encontrar el modelo matemático-estadístico que describa el patrón de crecimiento de frutos de naranja. Se registraron datos del diámetro ecuatorial (mm) de frutos en la provincia Entre Ríos, Argentina. Los datos experimentales se obtuvieron en 2 huertos comerciales de naranja 'Valencia late' (*Citrus sinensis* L. Osbeck) durante 2 temporadas. En cada huerto se seleccionaron 10 plantas, mediante un muestreo sistemático con arranque aleatorio. En cada planta se identificaron 30 frutos, distribuidos en toda la copa del árbol, a los que se realizó el seguimiento del diámetro ecuatorial (mm) empleando calibre digital, con una frecuencia 7 días. Las mediciones fueron referidas a días desde la fecha de plena floración (DDPF), que corresponde al 80% de flores abiertas. El mejor modelo no-lineal para describir el crecimiento de frutos de naranja Valencia Late, fue el logístico en su 5<sup>ta</sup> reparametrización. Como se espera que los parámetros del modelo varíen de acuerdo con las condiciones ambientales, un modelo mixto no lineal que incorpore efectos fijos y aleatorios es el adecuado para describir esta situación en particular. Con base en los datos experimentales, los modelos mixtos no lineales se ajustaron considerando los efectos fijos  $\beta_1$ ,  $\beta_2$ ,  $\beta_3$  y los efectos aleatorios  $b_1$ ,  $b_2$  y  $b_1$ - $b_2$ .-El proceso de ajuste fue para cada combinación de Temporada-Huerto y el ajuste de los parámetros se realizó con el método REML, y los efectos aleatorios de la covarianza: no estructurados, se realizaron con el software InfoStat, 2020. El valor AIC se utilizó como criterio de selección. El modelo que mejor se ajustó para la combinación Temporada-Huerto, fue el modelo con los dos parámetros aleatorios ( $b_1$ ,  $b_2$ ). Esto es consistente con los resultados encontrados para describir otras frutas ya que los parámetros  $\beta_1$ ,  $\beta_2$  están influenciados por las características ambientales (combinación Temporada-Huerto).

**Palabras clave:** *modelo logístico, efectos fijos y aleatorios, criterios de selección, frutales.*

### INTRODUCCIÓN

El crecimiento de frutos de naranjo dulce responde a una curva sigmoidea en forma de S, no lineal en sus parámetros. Para su ajuste se utilizan modelos de regresión no lineal. Las aplicaciones de modelos con supuestos simplistas para la componente del error no tienen en cuenta que en el caso de medidas longitudinales (repetidas en el tiempo sobre una misma unidad experimental), la estructura de la matriz de varianzas y covarianzas de los errores no se corresponde con tales asunciones. El hecho de realizar mediciones repetidas sobre el mismo sujeto implica que no es posible aleatorizar el factor tiempo, por lo que las medidas tomadas sobre un mismo individuo están autocorrelacionadas y, por lo tanto, no se cumple el supuesto

de independencia de los errores (Tassile, *et al.* 2002). El objetivo del presente trabajo fue evaluar efectos fijos y aleatorios de los parámetros en el modelo de crecimiento (Logístico en su 5<sup>ta</sup> reparametrización) de frutos de naranja 'Valencia late' seleccionado previamente.

## MATERIALES Y MÉTODOS

La toma de datos corresponde a la evolución de los diámetros desde 60 a 120 días después de plena floración hasta la cosecha, los cuales fueron obtenidos en huertos comerciales de naranja dulce (*Citrus sinensis* L. Osbeck) var. 'Valencia late', durante las temporadas 2006/07 (Temporada 1), 2007/08 (Temporada 2), en el departamento de Chajarí. Se obtuvieron entre 25 y 35 mediciones anuales repetidas del diámetro ecuatorial, para cada huerto comercial de diferentes edades (26-27 y 16-17 años) y temporada (2). El mejor modelo no-lineal para describir el crecimiento de frutos de naranja Valencia Late, fue el logístico en su 5<sup>ta</sup> reparametrización, (Avanza, *et al.* 2008). El proceso de ajuste fue para cada combinación de Temporada-Huerto. Se estimaron modelos considerando efectos fijos ( $\beta_1, \beta_2, \beta_3$ ) y aleatorios:  $b_1, b_2$ , y  $b_1-b_2$ . El ajuste de los parámetros se realizó con el método REML, y los efectos aleatorios de la covarianza: no estructurados, para la selección del modelo se utilizaron los criterios AIC y BIC. Los análisis de datos se realizaron con el software InfoStat, 2020.

$$DE (mm) = \frac{1}{\beta_1 + \exp(\beta_2)\gamma\beta_3^{DDPF}}$$

La interpretación de los parámetros en el modelo L5 según Ratkowsky (1983) es el siguiente:

DE (mm): diámetro ecuatorial del fruto

$\beta_1$ : parámetro relacionado con la asíntota superior (ASIN), en esta parametrización es la inversa ( $1 / \beta_1$ ).

$\beta_2$ : parámetro que relaciona la asíntota superior e inferior, describe el valor de Y correspondiente a DDPF = 0 (INT). Esta parametrización responde la función:  $[1 / (\beta_1 + e^{\beta_2})]$ .

$\beta_3$ : parámetro relacionado con la tasa de crecimiento (TC) desde los valores iniciales (magnitud  $\beta_2$ ) hasta los valores finales (magnitud  $\beta_1$ ), en esta parametrización está dado por la función  $-\ln \beta_3$ .

DDPF: días después de plena floración

## RESULTADOS Y DISCUSIÓN

En todos los casos analizados, los modelos no lineales mixtos tuvieron un ajuste mejor que los de efectos fijos y se seleccionó aquel modelo que contempló efecto aleatorio a los parámetros  $b_1$  y  $b_2$ , por presentar los menores valores de los criterios. AIC y BIC para cada combinación de huerto-temporada. Dentro de los modelos mixtos, aquellos con parámetros relacionado asíntota superior aleatoria ( $b_1$ ) y asíntota superior e inferior ( $b_2$ ) fueron los que mostraron el mejor ajuste en las combinaciones de huerto y temporada por lo que constituye una estrategia para modelar la covarianza entre las observaciones de los frutos. Los valores de los parámetros estimados se presentan en la tabla 1.

Tabla 1. Parámetros estimados del modelo logístico en su 5<sup>ta</sup> reparametrización por temporada y huerto en naranja Valencia late

Temporada	Huerto	Estimación		
		$b_1$	$b_2$	$b_3$
1	1	1,4559	-1,7983	0,9794

	2	1,5690	-1,8416	0,9798
2	1	1,5703	-2,5545	0,9810
	2	1,5464	-2,8766	0,9841

Las estimaciones de los parámetros  $b_1$  y  $b_3$  mostraron una mínima variación entre temporadas y huertos, en cambio la estimación de parámetro  $b_2$  se diferenciaron en mayor medida respecto de los huertos y temporadas. Por lo tanto la incorporación de efectos aleatorios en la modelización de la curva de crecimiento de frutos de naranja Valencia late, permite modelar la covarianza entre observaciones provenientes de un mismo fruto para contemplar la variabilidad existente entre individuos (frutos). Esto es consistente con los resultados encontrados para describir otras frutas ya que los parámetros  $\beta_1$ ,  $\beta_2$  están influenciados por las características ambientales (combinación Temporada-Huerto), es decir que se modifica el tamaño del fruto, tanto al inicio de crecimiento ( $\beta_2$ ) y su tamaño final ( $\beta_1$ ).

## REFERENCIAS

- [1] Avanza, M.M.; Bramardi, S.J.; & Mazza, S.M. 2008. Statistical models to describe the fruit growth pattern in sweet orange 'Valencia late'. Span J Agric Res. 6 (4), 577-585.2008.
- [2] Di Rienzo, J.A., Casanoves, F. Balzarini, MG., Gonzalez, L. Tablada, M., Robledo, CW. InfoStat, Universidad Nacional de Córdoba. 2011. Versión 2020 . Disponible en: Pagina web [www.infostat.com.ar](http://www.infostat.com.ar)
- [3] Ratkowsky, D.A. 1983, Nonlinear Regression Modeling: A Unified Practical Approach, New York: Marcel Dekker, Inc
- [4] Tassile, V.; Álvarez, O; Bramardi, S. 2002. Aplicación de modelos lineales mixtos en la descripción del crecimiento de frutos de pepita. V Congreso Latinoamericano de Sociedades de Estadística. Buenos Aires, Argentina. 133 p.

## VARIABILIDAD ESPACIO-TEMPORAL DE LA PRECIPITACIÓN EN EL LÍMITE OCCIDENTAL DE LA REGIÓN PAMPEANA. IMPLICANCIAS PARA LA AGRICULTURA

BOCA T.<sup>1</sup>; CIPRIOTTI P.A.<sup>1,2</sup>

<sup>1</sup> *Departamento de Métodos Cuantitativos y Sistemas de Información, Facultad de Agronomía, Universidad de Buenos Aires (FA-UBA).*

<sup>2</sup> *Instituto de Investigaciones Fisiológicas y Ecológicas Vinculadas a la Agricultura (IFEVA), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET).  
e-mail: boca@agro.uba.ar*

### RESUMEN

El cambio climático a escala global representa un desafío para muchos ecosistemas y la sustentabilidad de diversas actividades productivas, en especial las primarias. En general se caracteriza por ser un fenómeno muy heterogéneo a escala global y en particular en el límite occidental de la región Pampeana, donde se presenta la frontera agrícola para los cultivos en secano y una transición a la ganadería extensiva. Con el objetivo de contar con un modelo espacio-temporal de la precipitación que incluya el efecto del ENSO útil para la agricultura en esta zona de transición se estudió la precipitación asociada a los cultivos de verano (sep-mar), el índice MEI correspondiente al mismo período y los rendimientos de cultivos de verano comunes en la zona (maíz, sorgo y girasol) durante los últimos 50 años (1965-2014) para 67 localidades del este de La Pampa. Dada la complejidad de la estructura de los datos se aplicó un enfoque bayesiano utilizando la librería INLA de R. Se obtuvieron las estimaciones tanto de los efectos fijos como del campo aleatorio del modelo siguiendo un enfoque Bayesiano que permitió cuantificar la relación entre los cambios en la precipitación con los cambios en el MEI considerando la autocorrelación espacio-temporal. El valor medio para el coeficiente estimado asociado a la covariable MEI fue de 36,9 mm (25,9-47,9 mm). El rango de correlación espacial fue de aproximadamente 41 kilómetros y la asociación temporal entre años fue del 33%. Por otro lado, los coeficientes de asociación entre los valores de rendimiento y precipitación de verano diferenciados fueron significativos para los cultivos de maíz y sorgo, no así para el cultivo de girasol.

**Palabras clave:** *agricultura, cambio climático, INLA, modelos espacio temporales, precipitación.*

### INTRODUCCIÓN

El cambio climático a escala global representa un desafío para muchos ecosistemas que han visto alterados los regímenes de temperatura y precipitación de las últimas décadas amenazando la sustentabilidad de diversas actividades productivas y con agravamientos a futuro según las proyecciones de modelos atmosféricos. Está claro también que el calentamiento global es un fenómeno fuertemente heterogéneo a través de las distintas regiones del globo y en particular para la región Pampeana, sobre todo en el límite occidental de esta región donde existe una transición entre agricultura y ganadería. Según la Tercera Comunicación Nacional, en la región Pampeana argentina (Barros *et al.*, 2014) –a partir de 1960– las lluvias de primavera-verano aumentaron entre un 10% y un 50%. Se ha descrito

también una mayor ocurrencia de eventos extremos respecto de la precipitación, como sequías prolongadas o tormentas de alta intensidad, aumentando su variabilidad. En este contexto se ha producido un corrimiento de la frontera agrícola hacia el oeste de la Región Pampeana donde se presenta una mayor variabilidad inter-anual en las precipitaciones con un claro impacto sobre el rendimiento y la seguridad de cosecha de los principales cultivos.

De La Casa (2012) analizó la estructura de correlación para el conjunto disponible de indicadores climáticos del ENSO de enero a septiembre desde 1959 al 2008 y determinó que el MEI era uno de los indicadores que mostraba valores de correlación más altos con el total de lluvias del semestre octubre a marzo y su señal fue la más generalizada en toda la Provincia de Córdoba respecto de otros indicadores. Por otro lado Podestá *et al.* (1999), encontraron correlaciones significativas y similares entre las anomalías de las precipitaciones de noviembre a enero con las anomalías de los rendimientos de cultivos de verano, y determinaron que la precipitación de finales de la primavera a principios del verano parece mediar entre la asociación entre la fase ENSO y los rendimientos de maíz, sorgo y soja en la Pampa, aunque la correlación entre la precipitación y el rendimiento es menor en el sorgo que en el maíz, ya que generalmente el primero es más resistente al estrés hídrico. Maenza (2017) cubrió una región más amplia que los autores anteriores a través de un enfoque meteorológico y realizó un análisis de regionalización de las precipitaciones durante la estación húmeda sugiriendo cinco subregiones con precipitaciones espacialmente homogéneas, sus resultados mostraron la presencia de periodicidades a nivel interanual, decenal e interdecadal en las proximidades del noreste de la región bajo estudio, lo que sugiere posible conexión con El fenómeno Niño-oscilación sur (ENSO). A partir de los trabajos mencionados surge preguntarse sobre el modelo subyacente espacio - temporal de la precipitación y su relación con los indicadores del ENSO en el contexto de la agricultura extensiva en secano. Se consideró en particular los rendimientos de cultivos de verano: maíz, sorgo y girasol, la precipitación acumulada (desde septiembre hasta marzo) y los valores de MEI correspondiente a las campañas de los cultivos mencionados. El objetivo fue describir y modelar la variabilidad espacio-temporal de las precipitaciones en dicha zona, que ocupa más de 130.000 km<sup>2</sup>, su relación con el fenómeno ENSO y su relación con el rendimiento de los principales cultivos.

## MATERIALES Y MÉTODOS

**Precipitación:** Se generó una base de datos con los registros de precipitación mensual a partir de 67 sitios para el período 1965-2014 compaginados a partir de la información publicada por la red de la Administración Provincial del Agua de la pcia. de La Pampa (APA). El requisito para utilizar esos 67 sitios obedeció a contar con información completa para todos los meses durante el periodo informado.

**Índice del ENSO Multivariado (MEI):** este índice considera seis variables incluyendo la presión atmosférica al nivel del mar, los componentes zonales y meridionales del viento superficial, la temperatura superficial del mar, la temperatura del aire superficial y la fracción del cielo cubierto por las nubes en la zona tropical del pacífico. Según la GGWeather se definieron los eventos de El Niño (cálido) y La Niña (frío) en el Pacífico tropical según la metodología adoptada por Trenberth (1997).

**Análisis estadísticos:** Dada la complejidad de la estructura de los datos se aplicó un enfoque bayesiano. La principal ventaja de este enfoque reside en que tiene en cuenta incertidumbre en las estimaciones y su flexibilidad y capacidad para tratar con problemas como datos faltantes. Se ajustó un modelo con estructura de correlación espacio temporal (Blangiardo & Cameletti, 2013, 2015) utilizando inferencia bayesiana y la aproximación de Laplace anidada integrada

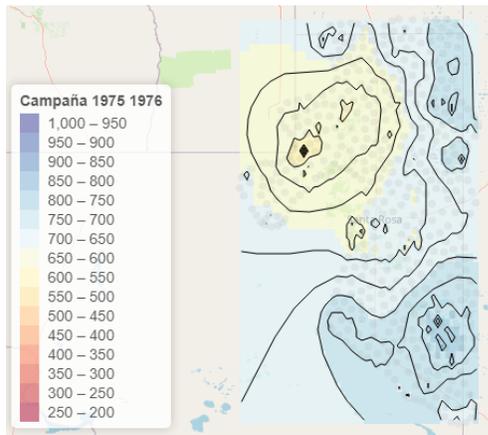
(INLA) a través del paquete R-INLA que ofrece una interfaz adecuada para el análisis de este tipo de datos.

Rendimientos históricos y precipitación: Se relacionaron los valores diferenciados de las series de precipitación con los rendimientos históricos a nivel departamental de los cultivos de verano (maíz, girasol y sorgo) para el periodo 1969/70-2014/15 obtenidos del Ministerio de Agricultura de la Nación.

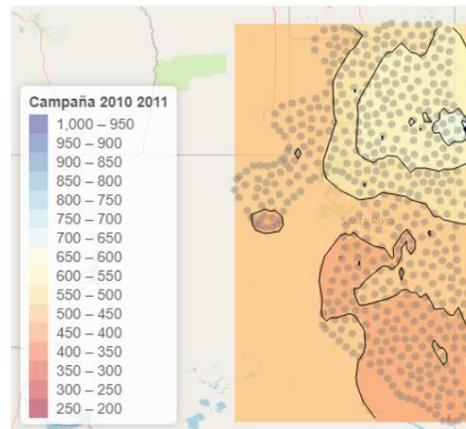
## RESULTADOS Y DISCUSIÓN

Se obtuvieron los estimadores puntuales de los efectos fijos y del campo aleatorio espacio-temporal del modelo mediante un enfoque Bayesiano que relacionó los cambios en la precipitación de la campaña de verano con los cambios en el MEI y que capturó tanto la autocorrelación espacial como la temporal. El valor medio para el coeficiente estimado asociado a la covariable MEI fue de 36,91 mm (25,86-47,95 mm). El rango de correlación espacial fue de aproximadamente 41 kilómetros y la asociación temporal fue del 33%. Por otro lado, los coeficientes de asociación entre los valores de rendimiento y precipitación de verano diferenciados fueron significativos para el cultivo de maíz y de sorgo, no así para el cultivo de girasol. A nivel fenómeno NIÑA-NIÑO los mapas obtenidos no muestran una relación unívoca según lo esperado siguiendo la definición del evento en el calendario de GGWeather.

1975/1976 Niña



2010/2011 Niña



La flexibilidad de esta aproximación nos permitirá seguir trabajando y realizar ajustes al modelo incluyendo otros indicadores climáticos del ENSO, nuevas covariables predictoras a los efectos fijos, nuevas modelizaciones del campo espacio-temporal, y probar alternativas para los algoritmos de ajuste mediante la especificación de nuevas distribuciones *a priori* de los hiperparámetros.

## REFERENCIAS

- [1] Blangiardo, M., & Cameletti, M. (2015). Spatial and spatio-temporal Bayesian models with R-INLA. John Wiley & Sons.
- [2] De La Casa, A. C. (2012, April). Capacidad predictiva del niño-oscilación sur (ENSO) y de otros indicadores climáticos sobre la precipitación en Córdoba, Argentina. In ANALES AFA (Vol. 22, No. 1).



- [3] Lindgren, F., Lindström, J., & Rue, H. (2010). An explicit link between Gaussian fields and Gaussian Markov random fields: The SPDE approach. Mathematical Statistics, Centre for Mathematical Sciences, Faculty of Engineering, Lund University.
- [4] Maenza, R. A., Agosta, E. A., & Bettolli, M. L. (2017). Climate change and precipitation variability over the western 'Pampas' in Argentina. *International Journal of Climatology*, 37, 445-463.
- [5] Podestá, G. P., Messina, C. D., Grondona, M. O., & Magrin, G. O. (1999). Associations between grain crop yields in central-eastern Argentina and El Niño–Southern Oscillation. *Journal of applied meteorology*, 38(10), 1488-1498.
- [6] Rue, H., Martino, S., & Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the royal statistical society: Series b (statistical methodology)*, 71(2), 319-392.
- [7] Trenberth, K. E. (1997). The definition of el nino. *Bulletin of the American Meteorological Society*, 78(12), 2771-2778.

## COMPARACIÓN DE MÉTODOS DE REGRESIÓN SIMBÓLICA PARA DATOS DE INTERVALO EN UN CONTEXTO BIOMÉTRICO

CARRIZO JORGELINA<sup>1</sup>, MALLEA ADRIANA<sup>1</sup>

<sup>1</sup> Departamento de Matemática, FFHA, Universidad Nacional de San Juan  
jcarrizo@ffha.unsj.edu.ar; jorgelinav\_carrizo@yahoo.com.ar; lamallea@ffha.unsj.edu.ar

### RESUMEN

La regresión, como herramienta estadística para la investigación de relaciones entre variables de respuesta y variables explicativas, ha sido durante mucho tiempo una de las técnicas centrales en el campo de la estadística. También se ha convertido cada vez más en un área de investigación activa en economía, psicología, educación, áreas bio-relacionadas, entre otras. La adaptación de la regresión a datos simbólicos, especialmente a datos con valores de intervalo, se ha transformado en un tema de creciente investigación desde que Billard y Diday (2000) introdujeron el primer enfoque para ajustar un modelo de regresión a datos con valores de intervalo. En el presente trabajo se presentan y aplican algunos métodos de regresión simbólica a datos obtenidos por agregación a partir de la matriz clásica “Hongos”, cuyas unidades son hongos extraídos del índice de especies de hongos de California y se caracterizan por las variables tipo de especie, ancho de la capa del píleo, longitud del estipe y grosor del estipe (Información más detallada sobre el conjunto de datos en [http://www.myknoweb.com/CAF/species\\_index.html](http://www.myknoweb.com/CAF/species_index.html).) Se comparan los distintos métodos con el error cuadrático medio, a fin de evaluar su performance.

**Palabras Claves:** *análisis simbólico de datos, variables simbólicas de intervalo, regresión simbólica, covarianza simbólica.*

### INTRODUCCIÓN

La regresión simbólica es parte del análisis simbólico de datos, se presentan adaptaciones del modelo de regresión lineal clásica para variables de intervalo, con los parámetros de regresión estimados por el conocido procedimiento de mínimos cuadrados usando promedios, varianzas y covarianzas simbólicas.

La metodología empleada forma parte de lo que se denomina Regresión simbólica de datos, área de investigación activa desde que Billard y Diday [1] introdujeron el primer enfoque del modelo de regresión a datos con valores de intervalo, trabajando con un modelo de regresión clásica sobre los centros de los intervalos. Posteriormente propusieron estimaciones a partir de modelos independientes para los límites inferior y superior de los intervalos. Lima Neto y de Carvalho [9], siguiendo el enfoque de Billard y Diday, emplearon los centros y rangos de los intervalos para construir dos regresiones independientes.

Sin embargo, la predicción del límite inferior podría exceder el límite superior previsto. Para evitar esta situación, Lima Neto y de Carvalho [8] consideraron un modelo restringido que obliga a todos los parámetros del modelo de los rangos a ser positivos.

El pensamiento de reducir los intervalos hasta los puntos centrales y rangos, para establecer un modelo; puede no reflejar la variación interna de los datos de la mejor manera desde un punto de vista simbólico.

Interesados en encontrar una solución que pueda usar directamente datos de valor intervalo, Xu [11] y Billard [4], propusieron un método de covarianza simbólico (llamado método SCM), para realizar regresión de datos con valores de intervalo. La idea principal es reconstruir el estimador de mínimos cuadrados de un modelo de regresión de una manera que utiliza la covarianza muestral simbólica.

Es bien sabido que en regresión de datos clásica el estimador resultante de la técnica de mínimos cuadrados es idéntico al estimador de estimación de verosimilitud si sólo se supone que los residuos siguen una distribución normal con estructura de covarianza conocida.

Partiendo de la función de verosimilitud propuesta por Le-Rademacher y Billard [8] para una variable aleatoria univariada con valores de intervalo, establecemos una función de verosimilitud básica para datos con valores de intervalo, que se deriva para obtener estimadores de máxima verosimilitud para modelos de regresión lineal. Mostramos que el procedimiento de mínimos cuadrados una vez más está de acuerdo con la estimación de máxima verosimilitud incluso en la situación de datos simbólicos con valores de intervalo.

Si bien existen otros métodos de Regresión Simbólica en la literatura, se mencionan sólo estos por ser los que se aplican en el presente trabajo.

## MATERIAL Y MÉTODOS

Los modelos de regresión simbólica de intervalos que se aplican en este análisis son:

*Método del Centro:* los coeficientes del modelo se estiman aplicando el modelo clásico al punto medio de los intervalos.

*Método Mínimo Máximo:* ajusta dos modelos de regresión lineal, para los límites inferior y superior del intervalo.

*Método del Centro y del Rango:* ajusta dos modelos de regresión lineal, para el punto medio y el rango del intervalo.

*Método de Centros y Rangos Restringidos:* coeficientes del modelo de centros se estiman como en el modelo anterior y coeficientes del modelo de rango se estiman utilizando el algoritmo de Lawson y Hanson.

*Método de Covarianza Simbólico:* Emplea el modelo de regresión lineal centralizada (modelo con variables centradas). Los coeficientes de regresión se estiman mediante el método de mínimos cuadrados, utilizando la matriz de covarianza simbólica. El método refleja la variabilidad del intervalo, como la dependencia entre todas las variables de una manera integral.

*Método de verosimilitud simbólica para datos de intervalo:* Utiliza la función de probabilidad simbólica para datos con valores de intervalo para obtener estimadores de máxima verosimilitud de los parámetros del modelo y estimadores de parámetros generales como media, varianza y covarianza.

El conjunto de datos de hongos, extraído del índice de especies de hongos de California, contiene mediciones de tres características que refieren a estipe (descripción del tallo del hongo) y capa de pileo (sombbrero del hongo) de 100 especies de hongos, miembros del género *Agaricales*. Concretamente las variables del conjunto son el ancho de la capa de pileo, longitud y grosor del estipe, todas variables numéricas clásicas.

Los métodos presentados se aplican a la tabla simbólica de datos, construida tomando como concepto cada especie de hongo, lo que da origen a tres variables simbólicas de tipo intervalo, obtenidas por agregación de las variables observadas en el conjunto de datos clásicos. Para los métodos de regresión simbólica se toma como respuesta el ancho de la capa del pileo y longitud del estipe como variable explicativa.

*Rendimiento de los modelos* (Lima Neto y De Carvalho [8]): La evaluación del desempeño de los modelos de regresión lineal descriptos, se basan en el error cuadrático medio para límites inferiores y superiores, que miden las diferencias entre los valores predichos y los valores observados:

$$RMSE_L = \sqrt{\frac{\sum_{j=1}^m (\underline{I}_{Y(j)} - \underline{I}_{\hat{Y}(j)})^2}{m}} \quad RMSE_U = \sqrt{\frac{\sum_{j=1}^m (\bar{I}_{Y(j)} - \bar{I}_{\hat{Y}(j)})^2}{m}}$$

## RESULTADOS

Para comparar los distintos métodos se ha utilizado, dentro de la plataforma R, el paquete “iRegression”. Este paquete tiene como autores a Lima Neto E., De Souza Filho C. y Marinho P. Contiene algunos métodos de regresión para variables simbólicas de intervalo. La Tabla1, resume la performance de los modelos, basada en sus respectivos errores cuadráticos medios.

Tabla 1: Evaluación del rendimiento de los modelos

Modelos	RMSE <sub>L</sub>	RMSE <sub>U</sub>
CM	1.89	5.02
MinMax	1.39	4.08
CRM	1.58	4,91
CCRM	1.56	4.90
SCM	1.82	4.99

Se observa que el método MinMax funciona mejor que los restantes métodos al estimar los coeficientes del límite inferior y del límite superior ya que proporciona el menor valor de RMSEL y menor valor RMSEU.

$$\hat{Y}_{\min} = 1,64 + 1,82\hat{X}_2 \min + \hat{e}_{\min} \quad \hat{Y}_{\max} = 3,60 + 2,69\hat{X}_2 \max + \hat{e}_{\max}$$

## CONCLUSIONES

En este trabajo se han presentado algunos modelos de regresión lineal simbólica para variables de intervalo, los cuáles han sido probados usando datos de clases de hongos, obtenidas por la especie a la que pertenecen. Se ha mostrado las ventajas de trabajar con regresión simbólica en el caso de tener clases de individuos en lugar de entes individuales y por ende intervalos de valores tanto para las variables explicativas como de respuesta.

La extensión de los supuestos probabilísticos, presentes en la teoría del modelo de regresión lineal clásica, al caso de datos de intervalo, sigue siendo un tema de análisis en la teoría de datos simbólicos que requiere más investigación. Esto motiva que a futuro se desarrollen metodologías de regresión simbólica, utilizando la aritmética de intervalos y respuestas de naturaleza aleatoria.

## REFERENCIAS

- [1] Billard and Diday (2000). Regression analysis for interval-valued data. In: Proc. of IFCS00, Belgium, pp. 369-374, Springer.
- [2] Billard, L., Diday, E. (2002). Symbolic Regression Analysis. In: Proc. IFCS02, Poland.
- [3] Billard, L., Diday, E. (2006). Symbolic Data Analysis: Conceptual Statistics and Data Mining.
- [4] Billard (2007, 2008). Sample covariance functions for complex quantitative data.
- [5] Bock, H.-H.; Diday, E. (2000): Analysis of Symbolic Data: Exploratory methods for extracting statistical information from complex data. Berlin-Heidelberg: Springer-Verlag.
- [6] De Carvalho, F.A.T., Neto, E.A.L. (2008). Centre and Range method for fitting a linear regression model to symbolic intervalar data. Computational Statistics, Data Analysis.
- [7] De Carvalho, F.A.T., Neto, E.A.L., (2010). Centre and Range method for fitting a linear regression model to symbolic intervalar data. Computational Statistics, Data Analysis.
- [8] Le-Rademacher, J. and Billard, L. (2010). Likelihood Functions and Some Maximum Likelihood.
- [9] Lima Neto, E. and F. De Carvalho (2008). Centre and Range method for fitting a linear regression model to symbolic interval data. Computational Statistics and Data Analysis.
- [10] Lima Neto, E. and F. De Carvalho (2010). Constrained linear regression models for symbolic interval-valued variables. Computational Statistics and Data Analysis 54(2), 333347. Estimators for Symbolic Data. Journal of Statistical Planning and Inference, Submitted.
- [11] Xu, W. (2010), Symbolic data analysis: regression of data with interval values.

## UN ANÁLISIS AUTOMÁTICO DE DATOS DE SOBREVIDA MEDIANTE R

KURISS CLAUDIA CASTRO<sup>1</sup>

<sup>1</sup>Instituto Tecnológico de Buenos Aires (ITBA), Departamento de Matemática.  
ccastro@itba.edu.ar

### RESUMEN

El comienzo de un análisis estadístico consiste en examinar las variables disponibles en un conjunto de datos, a qué tipo pertenece cada una, sus valores observados y si alguna presenta datos con características particulares como por ejemplo datos atípicos y faltantes. Con esta caracterización se visualizará cada variable y las relaciones entre ellas. La comprensión de los datos disponibles es fundamental para un modelado adecuado. En la actualidad, los datos son complejos por lo que resulta novedoso contar con herramientas que agilicen y automaticen esta caracterización, simplificando el análisis exploratorio. R es un entorno y lenguaje estadístico gratuito, compuesto de librerías que se pueden usar o crear para análisis de datos. Últimamente se han propuesto varias librerías en R de lo que se denomina “autoEDA” es decir análisis exploratorio de datos (EDA por sus siglas en inglés) automático.

El análisis de sobrevida representa un vasto campo de estudio en medicina e ingeniería. En esta área no se dispone aún de un análisis automático para los distintos tipos de datos con censura que permita además presentarlos de manera visualmente atractiva.

Los objetivos de este trabajo son (i) comparar paquetes que efectúan autoEDA para detectar sus diferencias y su utilidad aplicándolos a conjuntos de datos reales con censura progresiva y (ii) presentar un paquete similar en el tema de sobrevida, que se encuentra en desarrollo por el autor de este trabajo.

**Palabras clave:** *lenguaje R, análisis de sobrevida, autoEDA.*

### INTRODUCCIÓN

Los datos de sobrevida aparecen en ciencias de la salud y confiabilidad cuando no todas las piezas o pacientes bajo estudio experimentan la falla o el evento de interés durante el tiempo de seguimiento. Estos datos se denominan “censurados”. La variable para analizar es el tiempo hasta la ocurrencia del evento. Esta variable continua se suele explicar con covariables (continuas o categóricas) mediante el modelo de riesgos proporcionales (PHM), ver [3]. Cuando las suposiciones del PHM no se satisfacen, existen extensiones a este modelo con covariables que dependen del tiempo o con estratificación. El paquete principal para el análisis de sobrevida se llama **survival**, debido a Therneau [11] que junto con el paquete *KMsurv*, pueden aplicarse para analizar los datos de sobrevida en [4] pero es necesario recurrir a otras librerías para resultados complementarios. El tema se encuentra tratado en por lo menos 10 librerías (como por ejemplo en los paquetes R llamados *survminer*, *OLsurv*, *smoothSurv*). En particular, el análisis de residuos en el modelo de PHM está tratado en varias librerías lo que posibilita estudiar todos los tipos de residuos que se han definido (martingala, deviance, defbeta, Schoenfeld, Schoenfeld escalados, Cox-Snell). La extensión del PHM a covariables dependientes del tiempo resulta compleja y es necesario conocer la estructura del proceso de conteo involucrada para su análisis en R, ver [11]. Una vez modificado el archivo según este tipo de procesos, se pueden introducir tantas covariables dependientes del tiempo como se considere necesario. La validación del PHM tiene gran relevancia y sin embargo es omitida

generalmente por autores de trabajos del tema, ya sea por la dificultad que conlleva, o porque no cuentan con el software adecuado para llevarlo a cabo. R [8] permite realizar análisis para validar hipótesis, supuestos y estudiar residuos. El inconveniente es que hay que investigar varias librerías distintas para poder efectuarlo exhaustivamente. Mediante R es posible también hacer ajustes de modelos paramétricos. Últimamente se ha observado con mayor frecuencia en estudios clínicos ajustes de este tipo a los datos ya que estos procedimientos no eran habituales en la bibliografía médica, ver [7] por ejemplo.

En este trabajo se propone comparar algunas librerías que efectúan autoEDA para detectar sus diferencias, y al mismo tiempo presentar las necesidades en el área de sobrevida. Partiendo de un paquete ya publicado, se establece la necesidad de reunir más funciones y comandos en una sola librería para abordar este problema en forma más sencilla y eficaz. El libro de Moore [6] abarca muchos temas de sobrevida de forma breve y clara, con los comandos en R, pero no se ha encontrado, al conocimiento de los autores, un libro que reúna en su totalidad todas las librerías, comandos y ejemplos necesarios mediante el empleo de R, siendo [2] más exhaustivo. El presente trabajo contribuye en cuanto al estudio y análisis de las librerías necesarias y a su vez con el desarrollo de una librería que resuma el tema de interés abordado.

## **MATERIALES Y MÉTODOS**

Recientemente, Staniak y Biecek [10] analizaron diversos paquetes que efectúan autoEDA en R, haciendo énfasis en las características necesarias para ello, por ejemplo, la comprensión y análisis de los datos, su modelado y posterior evaluación. Dentro de este marco, se efectuó un estudio comparativo de varias librerías que lo realizan: *dataMaid*, *dataExplorer*, *dlookr*, *inspectdf* y *smartEDA*. De ellos se detectaron las características más relevantes y aquellas consideradas de mayor utilidad para un primer análisis de los datos. Se compararon con las posibilidades de las librerías que efectúan sobrevida, siendo la más empleada y conocida, *survival* debida a Therneau [11]. Sin embargo, ésta no es completa y se requieren por lo menos de 18 de ellas para recorrer una gran parte de un análisis de sobrevida. Se consideran como punto de partida, las librerías *survival* y *ggfortify*. Esta última debido a su capacidad gráfica ya que extiende el paquete gráfico *ggplot2* para datos con censura. A las características que permiten estos dos paquetes, hay que sumarle la posibilidad de realizar análisis con PHM con o sin covariables dependientes del tiempo, chequear las hipótesis necesarias, realizar un análisis de residuos y chequear el riesgo proporcional para las variables propuestas en un PHM.

Como ejemplos ilustrativos se emplearon dos conjuntos de datos reales, uno de pacientes con mieloma [5] y otro con datos de pacientes con carcinoma hepatocelular [9]. Los ejemplos permitieron analizar las variables involucradas, ambos con datos con censura progresiva a derecha, y proponer un modelo para explicar la sobrevida de estos pacientes, empleando técnicas ágiles y visualmente atractivas para obtener los resultados. Al mismo tiempo, si existe más de un modelo propuesto que ajusta los datos, decidir cuál es el mejor. Para el caso que éstos no sean anidados, se eligió el criterio de Akaike [1], que también debe ser incorporado a la librería en desarrollo.

## **RESULTADOS Y DISCUSIÓN**

Se encontraron las características distintivas de cada librería de autoEDA estudiada, priorizando posibilidades gráficas, la modificación de la paleta de colores y otros parámetros que se encuentran por default en cada una, junto con la elaboración de reportes y salidas exportables en algún formato (html, pdf, Latex o Word). De acuerdo con esto, el más completo resulta ser *DataExplorer* y luego *dlookr*. El primero es el único que efectúa análisis de

componentes principales y ambos calculan correlaciones lineales. *dllokr* efectúa el test de Shapiro-Wilk para cada variable continua y transforma las variables para normalizarlas. Estas opciones no resultan relevantes en un estudio con datos de sobrevivida.

El paquete *DataMaid* es el que tiene un buen reporte automático y es el más sencillo de todos. Por esta característica, es el más recomendable para iniciar un estudio de un conjunto de datos, para identificar problemas y depurarlos. Así se analizaron los datos de pacientes de mieloma para los que se propuso un modelo de Cox con 2 covariables dependientes del tiempo que resultó el mejor entre los distintos modelos formulados. También se analizaron los datos de pacientes con carcinoma hepatocelular, que contienen 23 variables cuantitativas y 26 cualitativas, con 10.22% de datos faltantes, donde no se incluyó el tiempo hasta la muerte sino solamente una variable definida como 1 si el paciente murió y 0 si no murió, por lo que no se pudo ajustar un PHM sino solamente describir el conjunto en su totalidad. Si bien a la fecha no se ha podido concluir la elaboración completa de una librería con todas las exigencias que se han señalado, se ha podido avanzar y se han establecido cuáles son las características más importantes que se desea abarcar para completar este objetivo, partiendo de las que realizan autoEDA.

## REFERENCIAS

- [1] Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19: 716-723.
- [2] Castro-Kuriss, C. (2018). *Análisis de supervivencia mediante el empleo de R*. 10.13140/RG.2.2.12736.84483/2. [https://www.researchgate.net/publication/325390160\\_Analisis\\_de\\_Supervivencia\\_mediante\\_el\\_empleo\\_de\\_R](https://www.researchgate.net/publication/325390160_Analisis_de_Supervivencia_mediante_el_empleo_de_R)
- [3] Cox, D.R., (1984). *Analysis of Survival Data*. Chapman and Hall, New York.
- [4] Klein, J.P. y Moeschberger, M.L. (2003). *Survival Analysis. Techniques for Censored and Truncated Data*. Springer, New York.
- [5] [Krall, J.M., Uthoff, V.A., Harley, J. B. (1975). A step-up procedure for selecting Variables associated with survival, *Biometrics*, 31, 49-57
- [6] Moore, D. F. (2016). *Applied Survival Analysis Using R*. Springer, New York.
- [7] Paz-Bailey, G. et al. (2018). Persistence of zika virus in body fluids. *New England of Journal Medicine*, 379:1234-1243.
- [8] R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. [www.R-project.org](http://www.R-project.org)
- [9] Seoane Santos M. et al. (2015) A new cluster-based oversampling method for improving survival prediction of hepatocellular carcinoma patients. *Journal of biomedical informatics*, 58,49-59.
- [10] Staniak, M. y Biecek, P. (2019). The Landscape of R Packages for Automated Exploratory Data Analysis. arXiv:1904.02101. Cornell University, Nueva York, EEUU. Versión obtenida 24/08/2019. <https://arxiv.org/abs/1904.02101>
- [11] Therneau, T. y Grambsch, P. (2000) *Modeling Survival Data: Extending the Cox Model*. Springer, New York.

## **ENSEÑANZA DE MATEMÁTICA Y BIOESTADÍSTICA EN MODALIDAD VIRTUAL**

CASTRO MJ<sup>1</sup>, NÚÑEZ M<sup>1</sup>, PONTEVILLE CH<sup>1</sup>

<sup>1</sup>*Cátedra Matemática, Facultad de Farmacia y Bioquímica, Universidad de Buenos Aires  
mjcastro@ffyb.uba.ar*

### **RESUMEN**

En el campo de las carreras de las Ciencias de la Salud, la enseñanza de la matemática y la bioestadística se encuentran en los fundamentos de la formación profesional. Dada las circunstancias del aislamiento social obligatorio, la cátedra de Matemática de la Facultad de Farmacia y Bioquímica de la Universidad de Buenos Aires implementó un formato virtual para el dictado de las asignaturas que en ella se imparten, durante el primer cuatrimestre de 2020. Con el fin de analizar y evaluar esta nueva modalidad, se les realizó a los estudiantes una encuesta no obligatoria, en la cual se abordaron diversas temáticas: experiencias respecto a la cursada virtual en sus diversas instancias, disponibilidades de recursos informáticos, participación en instancias sincrónicas y asincrónicas entre otras.

En este trabajo se propone un análisis sistemático de las actividades realizadas en relación con las respuestas dadas por los estudiantes, con el objetivo de identificar tendencias y elementos que permitan en un futuro modificar y/o incorporar diferentes modalidades y herramientas para la enseñanza de las áreas antes mencionadas.

Se clasificaron y resumieron los datos de la muestra mediante un análisis descriptivo y se analizó la independencia entre algunas de las variables estudiadas (Prueba de Chi-Cuadrado). El objetivo central de este trabajo consiste en obtener herramientas para analizar las condiciones didácticas en las que se desempeñaron los estudiantes bajo la modalidad virtual. Este análisis permite obtener información para tomar decisiones en próximas ediciones a realizarse de las asignaturas.

**Palabras clave:** *virtual, enseñanza, matemática, bioestadística, encuesta.*

### **INTRODUCCIÓN**

En la actualidad, las sociedades incorporan a la matemática y a la estadística entre sus prácticas de uso en la fundamentación de sus áreas de conocimiento, siendo su desarrollo y difusión una demanda cada vez más urgente. Los conceptos de alfabetización científica se encuentran en la base de la construcción del conocimiento en las áreas de las Ciencias de la Salud. Respecto del área de la matemática, se encuentra relacionada con los procesos de construcción de argumentaciones y de elementos básicos vinculados a la geometría, el cálculo infinitesimal, la estructura de procesamiento de información. En cuanto a la bioestadística, se construye la capacidad para poder explicar y valorar en forma crítica la información que se recibe respecto de datos o situaciones estocásticas que las personas pueden encontrar en diversos contextos, y la capacidad para discutir o participar sus opiniones respecto de estas ideas estadísticas cuando sea pertinente (Gal, 2002). Su desarrollo ha sido vertiginoso gracias a los aportes de la tecnología, de la ciencia y de los medios de comunicación.

Las Ciencias de la Salud no quedan fuera de este proceso social y existe una inquietud genuina de los docentes respecto a dar respuesta respecto de la diferencia que existe entre lo que es

enseñado y lo que es aprendido (Ponteville, 2012). Esta visión se sustenta en que los criterios de validación del conocimiento científico se establecen y por lo tanto son una construcción sociocultural que evoluciona y cambia no sólo con el tiempo sino de un grupo a otro.

Dadas las circunstancias del aislamiento social obligatorio, la Cátedra de Matemática de la Facultad de Farmacia y Bioquímica de la Universidad de Buenos Aires implementó un formato virtual para el dictado de las asignaturas a su cargo en el primer cuatrimestre de 2020. Con el fin de entender y perfilar las formas en las que se desarrollaron dichas cursadas se planteó a los estudiantes una encuesta en la cual se abordaron diversas temáticas: experiencias respecto a la cursada virtual en sus diversas instancias, disponibilidades de recursos informáticos, participación en instancias sincrónicas y asincrónicas entre otras.

En este trabajo se propone un análisis sistemático de lo planteado por los alumnos, de las actividades realizadas, con el objetivo de identificar tendencias y elementos que permitan en un futuro mejorar la enseñanza de esta área de formación.

## **MATERIALES Y MÉTODOS**

Se realizó una encuesta a los estudiantes de las asignaturas Matemática (433) y Bioestadística (204) de las carreras de Farmacia y Bioquímica; los de Matemática Aplicada (61) de la Tecnicatura Universitaria en Óptica Técnica y Contactología y los de Elementos de Matemática y Bioestadística (52) de la Tecnicatura Universitaria en Medicina Nuclear de la Facultad de Farmacia y Bioquímica de la Universidad de Buenos Aires.

Dicha encuesta fue planteada a través del Campus Virtual de las correspondientes asignaturas, la misma fue organizada a través de los siguientes conceptos: experiencia general en la cursada, disponibilidad y calidad de recursos informáticos, materiales disponibles en las diferentes instancias, características de encuentros sincrónicos y asincrónicos, formato de evaluación planteada.

Se clasificaron y resumieron los datos de la muestra mediante un análisis descriptivo y se analizó la independencia entre algunas de las variables estudiadas (Prueba de Chi-Cuadrado).

## **RESULTADOS Y DISCUSIÓN**

Para la asignatura Elementos de Matemática y Bioestadística de la Tecnicatura Universitaria en Medicina Nuclear se obtuvieron los siguientes resultados: el 65,4% de los estudiantes cursaron dicha asignatura por primera vez, el 46,2% cursó 3 materias o más. El 58,9% expresó que la experiencia sobre la cursada fue Buena - Muy Buena - Excelente, en cuanto al rendimiento el 38,5% respondió que fue Malo o Regular, mientras que para el 61,5% fue Bueno-Muy Bueno-Excelente. El 80,8% dispuso de más de un recurso informático y el 88,5% tenía servicio de internet. En relación con el material proporcionado por la cátedra el 78,8% estuvo Satisfecho o Muy Satisfecho. El 63,5% de los estudiantes contestó que prefiere las clases presenciales a las virtuales, mientras que al 21,1% las modalidades le resultaban equivalentes. En referencia al grado de satisfacción de las actividades sincrónicas el 69,2% manifestó estar Satisfecho o Muy satisfecho.

Para la asignatura Matemática Aplicada de la Tecnicatura Universitaria en Óptica y Contactología se obtuvieron los siguientes resultados: el 80,3% de los estudiantes cursaron dicha asignatura por primera vez, el 39,3% cursó 3 materias o más. El 98,4% expresó que la experiencia sobre la cursada fue Buena - Muy Buena - Excelente, en cuanto al rendimiento el 9,8% respondió que fue Malo o Regular, mientras que para el 90,2% fue Bueno-Muy Bueno-Excelente. El 90,2% dispuso de más de un recurso informático y el 98,4% tenía servicio de internet. En relación con el material proporcionado por la cátedra el 96,7% estuvo Satisfecho o

Muy Satisfecho. El 55,7% de los estudiantes contestó que prefiere las clases presenciales a las virtuales, mientras que al 22,9% las modalidades le resultaban equivalentes.

Para la asignatura Matemática correspondiente a las carreras de Farmacia y Bioquímica se obtuvieron los siguientes resultados: el 76,9% de los estudiantes cursaron dicha asignatura por primera vez, el 64,4% cursó 3 materias o más. El 89,6% expresó que la experiencia sobre la cursada fue Buena - Muy Buena - Excelente, en cuanto al rendimiento el 21,5% respondió que fue Malo o Regular, mientras que para el 78,5% fue Bueno-Muy Bueno-Excelente. El 92,4% dispuso de más de un recurso informático y el 98,4% tenía servicio de internet. En relación con el material proporcionado por la cátedra el 78,5% estuvo Satisfecho o Muy Satisfecho. El 68,8% de los estudiantes contestó que prefiere las clases presenciales a las virtuales, mientras que al 13,4% las modalidades le resultaban equivalentes.

Para la asignatura Bioestadística correspondiente a las carreras de Farmacia y Bioquímica se obtuvieron los siguientes resultados: el 74% de los estudiantes cursaron dicha asignatura por primera vez, el 49% cursó 3 materias o más. El 88,7% expresó que la experiencia sobre la cursada fue Buena - Muy Buena - Excelente, en cuanto al rendimiento el 20,1% respondió que fue Malo o Regular, mientras que para el 79,9% fue Bueno-Muy Bueno-Excelente. El 94,1% dispuso de más de un recurso informático y el 96,6% tenía servicio de internet. En relación con el material proporcionado por la cátedra el 83,3% estuvo Satisfecho o Muy Satisfecho. El 61,8% de los estudiantes contestó que prefiere las clases presenciales a las virtuales, mientras que al 14,7% las modalidades le resultaban equivalentes.

Si consideramos a los alumnos de las Tecnicaturas, la evaluación que realizan de su experiencia depende si cursa o no por primera vez la materia ( $P=0,0043$ ), en cambio en los alumnos de las carreras de grado es independiente ( $P=0,8824$ ). Existe asociación entre la valoración dada por el alumno de su rendimiento en la cursada en modalidad virtual y la asignatura ( $P=0,0028$ ). La preferencia sobre la modalidad virtualidad y/o presencial es independiente de la asignatura cursada ( $P=0,1419$ ).

## **REFERENCIAS**

- [1] Castro, M.J.; Bonadías, M.E.; Nuñez, M.; Ponteville, C. (2019). "Construcción de Conocimiento en Matemática y Bioestadística". Tucumán. República Argentina. XLVII Coloquio Argentino de Estadística.
- [2] Gal, I. (2020). "Adults' Statistical Literacy: Meanings, Components, Responsibilities". *International Statistical Review*, 70(1): 1-25.
- [3] Ponteville, Ch. (2012). "Reflexiones sobre la enseñanza de la estadística". *Premisa*, 14(54): 37-40.

## RALEO QUÍMICO, CAÍDA DE FRUTOS Y RETORNO FLORAL EN MANZANO RED CHIEF

CURETTI MARIELA<sup>1,2</sup>, RAFFO DOLORES<sup>1</sup>, ROMA FERNANDO<sup>1</sup>, RAGGIO FERNANDO, CASTRO ANDREA<sup>1</sup>, CORIA LAUTARO, REEB PABLO<sup>2</sup>

<sup>1</sup> INTA Estación Experimental Agropecuaria Alto Valle

<sup>2</sup> Dpto. de Estadística, Facultad Cs. Agrarias, Univ. Nacional del Comahue, Argentina  
curetti.mariela@inta.gob.ar

### RESUMEN

La eliminación temprana de una proporción de los frutos en manzanos mejora la calidad de los frutos restantes a cosecha y asegura un adecuado retorno a floración. Las giberelinas presentes en las semillas de los pequeños frutos en desarrollo inhiben la inducción floral, proceso que ocurre entre tres y ocho semanas luego de la floración. El objetivo de este trabajo ha sido describir la caída de frutos durante la primavera según distintos tratamientos de raleo químico aplicados y su relación con el retorno a flor en manzanos cv. *Red chief*. El número de frutos cosechados por árbol presentó una fuerte correlación negativa con el retorno a floración. Entre los tratamientos de raleo químico realizados se destacaron las aplicaciones de Metamitrona (MM) a 200 ppm cuando los frutos tenían 8 a 16 mm de diámetro. Estos tratamientos presentaron un mayor porcentaje de frutos caídos a los 15 - 21 días de la aplicación. El tratamiento de MM a 16 mm presentó un máximo retorno a floración en la temporada 2018/19 y no se diferenciaba aún del testigo a los 30 días después de la floración, lo que estaría indicando que la inducción floral en este cultivar aconteció posteriormente.

**Palabras clave:** *carga frutal, añerismo, inducción floral, momento de abscisión de frutos.*

### INTRODUCCIÓN

Los manzanos florecen en abundancia y tienden a retener un elevado número de frutos. La eliminación de una proporción de estos incrementa la relación hoja:fruto y mejora la calidad de los frutos a cosecha [1]. Otro propósito del raleo en manzano es asegurar un adecuado retorno a floración para controlar el “añerismo”, en el cual un año con elevada carga frutal es seguido por un año con muy baja carga. Entre los factores que explican este fenómeno se mencionan la menor disponibilidad de carbohidratos y un factor hormonal [2]. Las semillas en los pequeños frutos en desarrollo son fuente de giberelinas, las cuales inhiben la inducción de yemas florales en manzano. Este efecto inhibitorio producto de un exceso de frutos se sitúa entre las tres y ocho semanas luego de la floración [3]. Es por esto que el raleo manual, realizado habitualmente entre ocho y diez semanas luego de la floración, no logra evitar este problema. A diferencia del raleo manual, el raleo químico logra una remoción más temprana de frutos y resulta más efectivo en incrementar el retorno a floración [3]. Sin embargo, los autores no conocen estudios acerca del patrón de caída de frutos registrado luego de la aplicación de diferentes raleadores químicos, el cual podría influenciar el retorno a floración obtenido. El objetivo del presente trabajo ha sido describir la caída de frutos durante la primavera según distintos tratamientos de raleo químico y su relación con el retorno a flor.

## MATERIALES Y MÉTODOS

Se trabajó en un monte de manzano cv. `Red chief` sobre portainjerto MM111 plantado en el 2000 a 4 x 1,5 m. ubicado en la estación experimental agropecuaria del INTA Alto Valle durante las temporadas 2018-19 y 2019-20. Se evaluaron tres raleadores químicos: Ácido naftalen acético ppm, Benciladenina 100 ppm y Metamitrona (MM) a 200 ppm [4]. Las aplicaciones se realizaron en cinco momentos: 4, 8, 12, 16 y 20 mm de diámetro de fruto [5], registrándose los días transcurridos después de plena floración (ddpf). Además de estos 15 tratamientos, se evaluaron un testigo absoluto, un raleo manual y un raleo químico tradicional con Carbaril a 8 mm. Cada tratamiento fue aplicado en cinco árboles homogéneos (unidades experimentales) para la primera temporada en un DBCA. Debido al bajo y heterogéneo retorno a flor, en la 2ª temporada se categorizaron los árboles según el mismo y sólo se contó con cuatro repeticiones por tratamiento. Para la evaluación de la caída de frutos se marcaron seis ramilletes por árbol en floración y se realizó un seguimiento semanal del número de frutos persistentes durante los primeros 75 ddpf ( $\pm 150$  frutos por tratamiento de raleo y  $\pm 300$  frutos para el testigo) [6]. El número de frutos caídos por árbol se determinó en base a la diferencia en el número de frutos persistentes entre fechas consecutivas. Se determinó el número de frutos cosechados y producción por árbol. En base a estos valores, se estimó el peso medio de fruto. En la primavera siguiente se evaluó el retorno a floración en base al número de ramilletes florales por árbol [2].

El análisis estadístico de los datos se realizó con los softwares R (versión 3.5.1) e InfoStat (2018). En primera instancia se realizó un análisis descriptivo multivariado con análisis de componentes principales (ACP). Para estudiar la caída de frutos a lo largo del tiempo se ajustó un modelo logit debido a la naturaleza de la variable respuesta (distribución Bernoulli: fruto persistente vs fruto caído). Por último se realizaron pruebas  $\chi^2$  para la distribución de los frutos según el momento de ocurrencia de caída y la distribución de ramilletes según el número de frutos presentes, comparando tratamientos de raleo químico relevantes respecto del testigo.

## RESULTADOS Y DISCUSIÓN

A partir del ACP (Figura 1) podemos observar que el número de frutos cosechados por árbol se encuentra correlacionado negativamente con el peso medio de fruto y el retorno a floración. Entre los tratamientos de raleo químico realizados se destacaron MM.8mm (15-16ddpf), MM.12mm (21-22ddpf) y MM.16mm (27-29ddpf). Se observó una importante caída (45-47% en el testigo) en los primeros 23-28 días (Tabla 1), correspondiente principalmente a flores no polinizadas [7]. La segunda caída inicia a las 3-4 semanas luego de la floración y se prolonga por alrededor de seis semanas [8]. Esta caída se debe a la competencia entre los frutos y depende del balance de carbohidratos de los árboles [7]. Los tratamientos con MM presentaron un mayor porcentaje de frutos caídos respecto al testigo a los 15-21 días luego de la aplicación (Tabla 1). Más del 70% de los frutos terminaron cayéndose [8].

Los modelos logit estimados de probabilidad de caída de frutos ( $\hat{\pi}_{caidos}$ ) se observan en la figura 2 y son los siguientes:

$$\text{Testigo} = \text{logit}(\hat{\pi}_{caidos}) = \log\left(\frac{\hat{\pi}_{caidos}}{1-\hat{\pi}_{caidos}}\right) = -2,31 + 0,0605 \cdot DDPF \quad \hat{\pi}_{caidos} = \frac{e^{-2,31+0,0605 \cdot DDPF}}{1 - e^{-2,31+0,0605 \cdot DDPF}}$$

$$\text{MM.16} = \text{logit}(\hat{\pi}_{caidos}) = \log\left(\frac{\hat{\pi}_{caidos}}{1-\hat{\pi}_{caidos}}\right) = -3,18 + 0,0956 \cdot DDPF \quad \hat{\pi}_{caidos} = \frac{e^{-3,18+0,0956 \cdot DDPF}}{1 - e^{-3,18+0,0956 \cdot DDPF}}$$

Es interesante destacar que el tratamiento de MM a 16 mm logró el máximo retorno a floración en la temporada 2018-19. A los 30 ddpf, la probabilidad de caída de fruto de estos tratamientos fue similar ( $0,40 \pm 0,02$ ); mientras que a los 45 ddpf esta probabilidad fue superior para el tratamiento con MM ( $0,75$  vs  $0,60$ ). Esto podría estar indicando que la inducción floral en este cultivar tuvo lugar luego de los 30 ddpf.

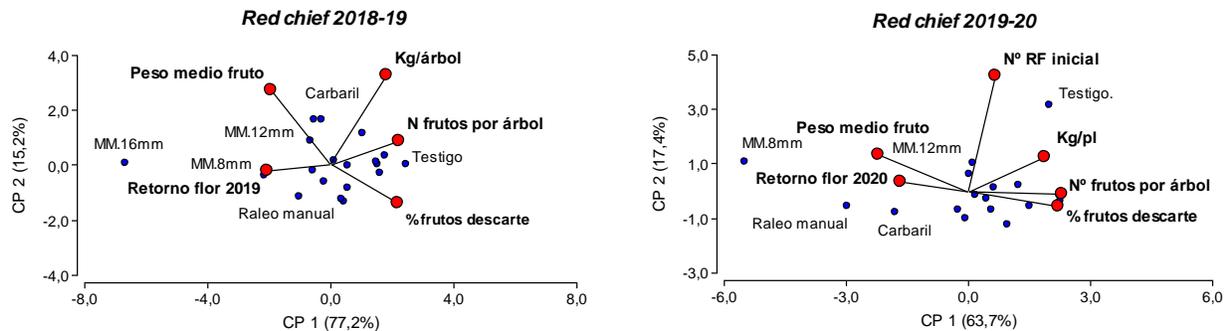


Figura 1. Biplot del Análisis de Componentes Principales realizado para los resultados de cosecha y retorno a floración en manzano Red chief.

Tabla 1. Distribución porcentual de los frutos según momento de ocurrencia de caída en días después de plena floración (ddpf) y aplicación de Metamitrona (MM) a 200 ppm.

Tratamiento	Momento de caída en 2018 (ddpf)						Persis- tentes	p-valor
	14-21	21-28	28-35	35-42	42-49	49-77		
Testigo	23,3	23,9	10,4	5,2	7,1	11,7	18,4	-
MM.16ddpf	25,1	26,3	18,6	10,2	1,8	2,4	15,6	0,0001
MM.21ddpf	33,3	12,6	7,4	20,7	9,6	2,2	14,1	<0,0001
MM.29ddpf	32,4	18,7	4,3	1,4	33,1	5,0	5,0	<0,0001

### Gráfico de efectos

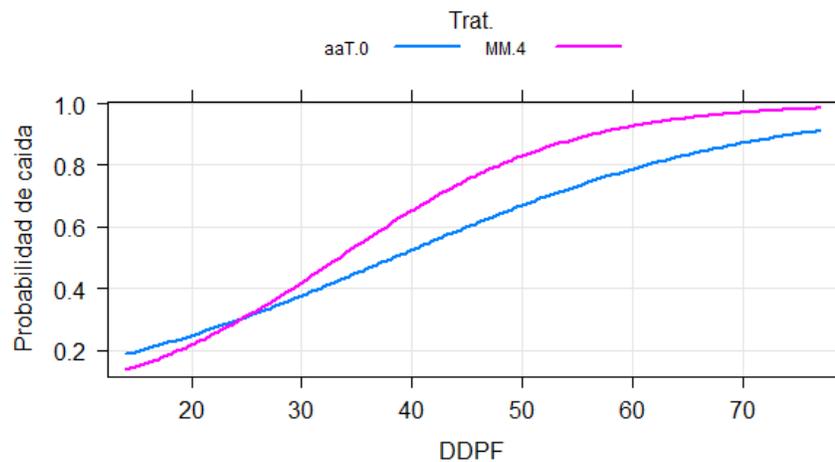


Figura 2. Gráfico de probabilidad de caída de los frutos de manzano Red chief para el testigo (en azul) y el tratamiento con MM a 200 ppm (en rosa) en el 4º momento de aplicación (29 ddpf-16 mm).

## REFERENCIAS

- [1] Wünsche, J.N. and Ferguson, I.B. (2005). "Crop load interactions in apple". *Horticultural reviews* vol (31): 231-290.
- [2] Meland, M. (2009). "Effects of different crop loads and thinning times on yield, fruit quality, and return bloom in *Malus domestica*". *J. Hort. Sci. and Biotech.* vol (84): 117-121.
- [3] Tromp, J. (2000). "Flower-bud formation in pome fruits as affected by fruit thinning". *Plant Growth Regulation* vol (31): 27-34.
- [4] Wertheim, S. J. (2000). "Developments in the chemical thinning of apple and pear". *Plant Growth Regulation* vol (31): 85-100.
- [5] Greene, D. W. (2002). "Chemicals, timing, and environmental factors involved in thinner efficacy on apple". *HortScience* vol (37): 477-481.
- [6] Striebeck, G., Dussi, M.C., Giménez, G., Gastiazoro, J., Barnes, N., Racskó, J, Nyeki, J. (2008). "Abscisión de Flores y Frutos en Manzanos (*Malus domestica* Borkh)". *Horticultura Argentina* vol (27).
- [7] Racskó J., Leite G.B., Petri, J.L., Zhongfu, S., Wang, Y., Szabó, Z., Soltész, M., Nyéki, J. (2007). "Fruit drop: The role of inner agents and environmental factors in the drop of flowers and fruits". *International J. Hort. Sci.* vol (13): 13-23.
- [8] Jakopic, J., Cebulj A., Schmitzer V. and Veberic R. (2018). "Fruitlet shedding is affected by the position in the apple cluster". *Acta Horticulturae* vol (1228): 323-330.

## IDENTIFICACIÓN DE CLONES DE BANANA (*MUSA SPP.*) PARA LA CREACIÓN DE UN BANCO DE GERMOPLASMA

DEL MEDICO ANA PAULA<sup>1</sup>, TENAGLIA GERARDO<sup>2</sup>, VITELLESCHI MARÍA SUSANA<sup>3</sup>,  
LAVALLE ANDREA<sup>4</sup>, PRATTA GUILLERMO RAÚL<sup>1</sup>

<sup>1</sup>*Instituto de Investigaciones en Ciencias Agrarias de Rosario (IICAR), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)*

<sup>2</sup>*Instituto Nacional de Tecnología Agropecuaria (INTA)*

<sup>3</sup>*Facultad de Ciencias Económicas y Estadística, Instituto de Investigaciones Teóricas y Aplicadas de la Escuela de Estadística, Universidad Nacional de Rosario (UNR)*

<sup>4</sup>*Departamento de Estadística, Universidad Nacional del Comahue (UNCOMA)*

*lavalleandrealina@gmail.com*

### RESUMEN

En un híbrido de reproducción asexual como banana (*Musa spp.*) la evaluación de clones a corto plazo se ve limitada por no disponer del número adecuado de repeticiones. Con el objetivo de caracterizar una población de clones de banana para establecer criterios para la conformación de un banco de germoplasma de banana, se aplicó la técnica de Análisis Factorial Múltiple Mixto (AFMmix) a una población compuesta por 124 clones de banana recolectados en diferentes campos de productores y por 4 testigos de uso comercial. Se construyó un índice de selección (IS) a partir de las coordenadas del AFMmix de forma de ordenar a los clones y poder seleccionar un subconjunto de los mismos que permita conservar la variabilidad genética existente. Se evaluaron 2 grupos de variables relacionados con la aptitud agronómica de los clones, uno conformado por 9 variables fenotípicas cuantitativas y el otro por 3 variables fenotípicas cualitativas. Los dos primeros ejes del AFMmix explicaron un 49,47% de la variabilidad total de los datos. Los atributos cuantitativos que más contribuyeron a la formación del primer eje fueron altura y diámetro de la planta, peso del raquis y peso de manos. Los caracteres cualitativos que más aportaron a dicho eje fueron tamaño de racimo y prolificidad de la mano. En el segundo eje no se observaron contribuciones considerables de las variables cuantitativas, sin embargo, sí lo hizo el carácter cualitativo foliosidad. En efecto, mediante la técnica de AFMmix se logró caracterizar al conjunto de clones de banana por los caracteres cuantitativos y cualitativos. Esto posibilitó identificar los clones de manera tal que permite determinar un subconjunto que presente la mayor diversidad teniendo en cuenta ambos tipos de caracteres.

**Palabras clave:** *análisis factorial múltiple mixto, índice de selección, análisis multivariado.*

### INTRODUCCIÓN

Un banco de germoplasma es una colección de material vegetal vivo que tiene como objeto conservar la variabilidad genética existente en una o varias especies de interés. Los bancos de germoplasma son el medio principal para proteger la fitodiversidad de las diferentes especies vegetales e identificar accesiones para los programas de mejoramiento.

La banana (*Musa spp.*) es un cultivo que tiene una importancia fundamental para las economías de muchos países en desarrollo. Siendo un híbrido de reproducción asexual, la selección de

clones, en el corto plazo, se ve limitada por no disponer del número adecuado de repeticiones. En el norte argentino, el cultivo de banana sufre de condiciones climáticas sub-óptimas, afectando a la diversidad. Por lo tanto, se presentan genotipos adaptados a ambientes menos favorables para la producción tradicional. Es así que en Argentina se presenta una diversidad del cultivo única en el mundo (Ermini *et al.*, 2018).

El objetivo de este trabajo fue proponer al AFMmix como una herramienta para caracterizar una población de clones de banana a fin de establecer, a largo plazo, criterios para la conformación de un banco de germoplasma de banana.

## MATERIALES Y MÉTODOS

La población bajo estudio estuvo compuesta por 124 clones de banana recolectados en diferentes campos de productores de la provincia de Formosa, Argentina (Ermini *et al.*, 2018) y por 4 testigos de uso comercial: las variedades comerciales Jaffa, Gal, Gran Enanao y Williams. Se realizó un “diseño en bloques aumentado” (Cotes & Núñez, 2001) de 15 bloques con 14 plantas cada uno, donde sólo los testigos poseían repeticiones. El mismo se llevó a cabo en las instalaciones de INTA Formosa, ubicadas en el norte argentino (26°11'31.8"S, 58°12'22.4"W), durante la campaña 2016-2017.

Se evaluaron 2 grupos de variables relacionados con la aptitud agronómica de los clones, uno conformado por 9 caracteres fenotípicos cuantitativos continuos y el otro por 3 caracteres cualitativos. Las variables cuantitativas fueron: altura de la planta (en m), diámetro de la planta (en cm), peso del raquis (en kg), peso de las manos (en kg), diámetro de la segunda mano (en cm), diámetro de la última mano (en cm), longitud de la segunda mano (en cm), longitud de la última mano (en cm) y grosor de cáscara (en mm). Los caracteres cualitativos fueron foliosidad (alta o baja), tamaño de racimo (grande o chico) y prolificidad de la mano (alta o baja).

Se aplicó la técnica AFMmix (Pagès, 2014). Esta metodología permite analizar simultáneamente dos grupos de variables de distinta naturaleza (cuantitativa y cualitativa), evaluados sobre el mismo conjunto de individuos. La razón de emplear AFMmix fue otorgar similar ponderación a ambos tipos de caracteres. Se construyó un índice de selección (SI) a partir de las coordenadas de las variables cuantitativas obtenidas mediante el AFMmix. Dicho índice consistió en ponderar a las variables estandarizadas, cuya contribución a cada factor superara los 2/3 del correspondiente coeficiente máximo en valor absoluto, por la inercia explicada por dicho factor. La construcción de este SI se basó en Del Medico *et al.* (2018), donde se aborda el desarrollo de un SI con el fin de seleccionar los mejores clones de banana para crear una variedad sintética.

## RESULTADOS Y DISCUSIÓN

Los dos primeros ejes del AFMmix explicaron un 49,47% de la variabilidad total de los datos. Los atributos cuantitativos que más contribuyeron a la formación del primer factor fueron peso de las manos, peso del raquis, diámetro y altura de la planta. Sobre el segundo eje no se observaron contribuciones considerables de las variables cuantitativas (Tabla 1). Los caracteres cualitativos ordinales que más aportaron al primer factor fueron tamaño de racimo y prolificidad de la mano. Sobre el segundo eje, la mayor contribución fue del carácter cualitativo foliosidad, seguido por tamaño de racimo (Tabla 2).

Se procedió a clasificar a los clones de banana según su foliosidad, dado que dicho carácter fue el que presentó mayor contribución sobre la construcción del segundo factor. Se formaron de esta manera 2 grupos de clones: uno compuesto por el testigo 3 y 46 clones correspondientes a plantas con baja foliosidad y el otro compuesto por los testigos 1, 2 y 4 y 78 clones

correspondientes a plantas con alta foliosidad. Luego, para seleccionar a los clones de banana según los caracteres cuantitativos que más aportaron a la construcción del primer factor, se construyó el SI de manera de ordenarlos dentro de cada uno de los grupos.

Tabla 1: Coordenadas de las variables cuantitativas sobre el primer plano factorial del Análisis Factorial Múltiple Mixto (AFMmix)

Variables	F1	F2
altura	0.62	-0.20
diámetro	0.77	0.01
peso del raquis	0.82	0.06
peso de manos	0.83	0.08
diámetro segunda mano	0.13	0.02
longitud segunda mano	0.52	0.18
diámetro última mano	0.14	0.22
longitud última mano	0.29	0.31
grosor de cascara	-0.12	0.03

Tabla 2: Coordenadas de las categorías de las variables cualitativas sobre los dos primeros ejes factoriales del Análisis Factorial Múltiple Mixto (AFMmix)

Variables	F1	F2
Foliosidadbaja	-0.56	-0.92
Foliosidad alta	0.32	0.53
Tamaño de racimo chico	-1.62	0.64
Tamaño de racimo grande	0.45	-0.18
Prolifricidad baja	-1.08	0.11
Prolifricidad alta	1.00	-0.10

Debido a que las variables cuantitativas no presentaron contribuciones considerables sobre el segundo factor de AFMmix (Tabla 1), se incluyó en la construcción del SI sólo el primer factor. En base a los datos presentados en la Tabla 1 y considerando que el primer autovalor del AFMmix fue igual a 1.64, el índice construido fue:

$$SI = 1.64 * (0.62 * altura + 0.77 * diámetro + 0.82 * peso raquis + 0.83 * peso manos)$$

Finalmente, se ordenaron los clones de banana de acuerdo a dicho índice dentro de cada uno de los dos grupos ya determinados. La selección en cada uno de los grupos fue proporcional a su tamaño. Para conservar la variabilidad genética existente se recomienda seleccionar, dentro de cada grupo, clones con SI alto, moderado y bajo.

Se concluye que mediante la técnica de AFMmix se logró caracterizar al conjunto de clones de banana por los caracteres cuantitativos y cualitativos. Esto permitió identificar los clones de forma de determinar un subconjunto que presente la mayor diversidad teniendo en cuenta ambos tipos de caracteres.

## REFERENCIAS

- [1] Del Medico, A. P., Ermini, J. L., Tenaglia, G., Vitelleschi, M. S., Lavallo, A., Pratta, G. R. (2018). Índice de selección para múltiples caracteres en una población de banana (*Musa spp.*). *Libro de Resúmenes XXIII Reunión Científica del Grupo Argentino de Biometría. II Encuentro Argentino - Chileno de Biometría*
- [2] Cotes, J. M., & Núñez, C. E. (2001). Propuesta para el Análisis de Diseños Aumentados en Fitomejoramiento: Un Caso en Papa. *Revista Latinoamericana de la Papa*, 12:15-34



- [3] Ermini, J. L., Tenaglia, G. C. & Pratta, G. R. (2018). Molecular diversity in selected banana clones (Musa AAA "Cavendish"): adapted to the subtropical environment of Formosa Province (Argentina). *American Journal of Plant Sciences*, 9(12): 2504-2513
- [4] Pagès, J. (2014). *Multiple factor analysis by example using R*. Chapman and Hall/CRC.

## APLICACIÓN DE MODELOS MIXTOS NO LINEALES EN EL CRECIMIENTO DE MANZANAS CV. 'RED CHIEF'

DEL BRIO DOLORES<sup>1,2</sup>, TASSILE VALENTÍN<sup>3</sup>, REEB PABLO D.<sup>2,4</sup>

<sup>1</sup> *Inst. Nacional de Tecnología Agropecuaria y Consejo Nacional de Investig. Científicas y Técnicas*

<sup>2</sup> *Dpto. de Estadística, Facultad de Ciencias Agrarias, Univ. Nac. del Comahue, Argentina*

<sup>3</sup> *Facultad de Ciencia y Tecnología de los Alimentos, Univ. Nac. del Comahue, Argentina*

<sup>4</sup> *Centro de Inv. en Toxicología Ambiental y Agrobiotecnología del Comahue, CONICET-UNCo  
delbrio.dolores@inta.gob.ar*

### RESUMEN

Las curvas de crecimiento de frutos de pepita pueden ser utilizadas para realizar pronósticos de producción. Su utilidad reside en que dichos pronósticos además de estimar cuantitativamente el rendimiento potencial, permiten estimar en parte la calidad de la fruta al separarla por tamaños comerciales. El objetivo de este trabajo consistió en proponer modelos mixtos no lineales para describir el crecimiento de frutos de manzana cv. "Red Chief" en el Alto Valle de Río Negro y Neuquén a lo largo de tres temporadas. Los modelos resultantes consideraron a las temporadas y la categoría de tamaño inicial como efectos fijos, a los frutos individualizados como efectos aleatorios y contemplaron la heterocedasticidad prevaleciente en las temporadas. Del análisis del modelo finalmente propuesto se pudo apreciar un claro efecto de categoría de tamaño inicial y temporada para todos los niveles de ambos efectos considerados.

**Palabras clave:** *manzanas, patrón de crecimiento, modelos mixtos.*

### INTRODUCCIÓN

El rendimiento a cosecha de un árbol depende del número de frutos que el mismo posee y del peso que los frutos alcanzan al momento de cosecha. Estimar el tamaño al que llegarán los frutos en fechas previas a la cosecha es de gran utilidad, ya que permite definir estrategias de manejo en las tareas de campo, empaque, conservación, y comercialización, a los efectos de maximizar el mercado diferencial de precio de la fruta según su tamaño [5]. Diferencias en algunos gramos en el peso de los frutos pueden representar grandes pérdidas en los retornos por exportación. Es por ello que la predicción de la distribución final de tamaños de la fruta es una práctica difundida en muchas regiones de producción de frutales de pepita. En los últimos años, los pronósticos de producción que utilizan curvas de crecimiento han demostrado ser muy útiles ya que, además de la predicción cuantitativa de la producción total potencial, pueden estimar en parte la calidad de la misma al separarla por tamaños comerciales [2].

Las curvas de crecimiento intentan describir el patrón de crecimiento de los frutos caracterizados por comportamientos no lineales del tipo sigmoideo. Este tipo de modelos se caracterizan por presentar una asíntota superior que indica el tamaño máximo que puede alcanzar el fruto, un punto de inflexión donde la velocidad del proceso comienza a decrecer y una asíntota inferior.

En frutales de pepita la curva logística en su tercera parametrización según Ratkowsky, D.A. [4], describe bien el crecimiento de los frutos [1], cuya propuesta permite, además, una adecuada interpretación de los parámetros del mismo en términos del crecimiento de los frutos. Es

necesario aclarar, sin embargo, que cada ajuste obtenido modela patrones de crecimiento para predecir tamaños para una variedad en una zona determinada, por lo cual no sería apropiado extrapolar dichos modelos en diferentes condiciones de cultivo.

El objetivo del presente trabajo es proponer modelos mixtos no lineales para describir el crecimiento de frutos de manzana “Red Chief” en el Alto Valle de Río Negro durante tres temporadas.

## MATERIALES Y MÉTODOS

Los datos obtenidos corresponden a una parcela de manzana cv. ‘Red Chief’ de la estación experimental INTA Alto Valle. Dicha parcela fue implantada en el año 2000 sobre portainjerto MM111 con una superficie de 0,8 ha. Posee riego por aspersión y recibe un manejo similar a los montes comerciales de la región.

Los datos fueron obtenidos durante las temporadas 2017-2018, 2018-2019 y 2019-2020. Al inicio de cada temporada de crecimiento se seleccionaron al azar 5 plantas de manzana. En cada una de ellas se seleccionaron 8 frutos chicos, 8 frutos medianos y 8 frutos grandes, resultando un total de 120 frutos. Semanalmente se obtuvo el promedio de dos mediciones opuestas del diámetro ecuatorial de los frutos utilizando un calibre digital (Essex 150 mm, Stainless Hardened). Los frutos caídos fueron reemplazados por un otro que tuviera un diámetro similar. Las mediciones comenzaron aproximadamente a los 30 días posteriores a plena floración (DDPF) y finalizaron una semana después de la cosecha.

Una vez digitalizados los datos, se procedió a realizar las estimaciones a través del paquete estadístico NLME del software estadístico R. Este obtiene la verosimilitud a partir de una expansión de Taylor de acuerdo al método de linealización BLUP propuesto por Lindstrom, M. L. & Bates, D. M. [3] para obtener las estimaciones de NLME con efectos aleatorios. Una vez ajustado el modelo, las componentes de varianza se evaluaron a partir de Criterios de Información y Razón de Verosimilitud para luego confirmarse, en lo posible, con intervalos de confianza obtenidos a partir del estadístico de Wald.

## RESULTADOS Y DISCUSIÓN

Para la modelación se contemplaron modelos alternativos obtenidos a partir de distintas propuestas que consideraban, para todos los parámetros, como efectos fijos a los factores temporada y categoría de tamaño comercial; y aleatorios en los niveles planta y fruto dentro de planta. El modelo final seleccionado tiene como efectos fijos a temporada y tamaño, efectos aleatorios a nivel de fruto y varianza residual que varía de acuerdo a cada temporada considerada. La expresión del modelo final es la siguiente:

$$Y_{jkl} = \frac{1}{(\beta_{1,TAM_k} + \beta_{1,TEMP_l} + b_{1,j}) * 0.01 + ((\beta_{2,TAM_k} + \beta_{2,TEMP_l} + b_{2,j}) * 0.01) * ((\beta_{3,TAM_k} + \beta_{3,TEMP_l} + b_{3,j}) * 0.1)^{DDPF} + \varepsilon_{jkl}}$$

donde:  $\beta_{i,TAM_k}$  es el efecto del k-ésimo tamaño para el i-ésimo parámetro del efecto fijo (i:  $\beta_1, \beta_2, \beta_3$ ; k: pequeño, mediano y grande);  $\beta_{i,TEMP_l}$  es el efecto de la l-ésima temporada para el i-ésimo parámetro del efecto fijo (i:  $\beta_1, \beta_2, \beta_3$ ; l: 2017-2018, 2018-2019, 2019-2020),  $b_{i,j}$  es el efecto aleatorio del j-ésimo fruto del i-ésimo parámetro con  $b \sim N(0,D)$ .

A continuación, se pueden apreciar las estimaciones por Máxima Verosimilitud restringida del modelo final propuesto y la inferencia correspondiente para los efectos fijos:

Efectos aleatorios:

	StdDev	Corr	
A (Intercept)	0,1484	A (ln)	B (ln)

B (Intercept)	1,2667	-0,375	
C (Intercept)	0,0278	0,082	-0,727
Residual	2,5798		

Estructura de varianza: parámetros estimados:

Temporada	2018-2019	2017-2018	2019-2020
	10.000	0,4889	0,6127

Efectos fijos: A + B + C ~ Temporada + Tamaño

	Value	Std.Error	t-value	p-value
A (Intercept)	1,3797	0,0166	82,85	0,0000
A Temporada1819	-0,0172	0,0192	-0,89	0,3715
A Temporada1920	-0,1029	0,0187	-5,48	0,0000
A Tamaño M	0,0886	0,1913	4,62	0,0000
A Tamaño P	0,2090	0,0192	10,91	0,0000
B (Intercept)	6,7315	0,1551	44,93	0,0000
B Temporada1819	-0,6511	0,1955	-3,41	0,0006
B Temporada1920	-0,8147	0,1729	-4,71	0,0000
B Tamaño M	0,9060	0,1765	5,13	0,0000
B Tamaño P	2,4710	0,1851	13,39	0,0000
C (Intercept)	9,6931	0,0035	2759,30	0,0000
C Temporada1819	-0,0067	0,0047	-1,43	0,1518
C Temporada1920	0,0128	0,0040	3,16	0,0016
C Tamaño M	-0,0136	0,0042	-3,16	0,0015
C Tamaño P	-0,5054	0,0043	-9,34	0,0000

Los resultados ponen en evidencia la necesidad de considerar efectos aleatorios a nivel de fruto para todos los parámetros considerados, la presencia de heterocedasticidad residual motivada por comportamientos disímiles para cada una de las temporadas de crecimiento y el claro efecto de la categoría de tamaño (pequeño, mediano y grande) sobre los calibres ecuatoriales obtenidos. La tasa de crecimiento y el tamaño final tienen comportamientos diferenciales con respecto al tamaño inicial en lo que respecta al efecto de la temporada de crecimiento. En la tasa de crecimiento y el tamaño final no se obtuvieron diferencias significativas entre las temporadas 2017-2018 y 2018-2019, mientras que para el tamaño inicial la falta de significación se halló entre las temporadas 2018-2019 y 2019-2020. Estos resultados sugieren la existencia de procesos diferenciales externos para los parámetros del crecimiento considerados por lo que se propone para un trabajo posterior la caracterización y cuantificación de fuentes de variabilidad ambientales que permitan explicar los hallazgos del presente trabajo.

## REFERENCIAS

[1] Bramardi, S.J., Castro, H. R. & Zanelli, M. L. (1997). Fruit growth pattern of pear cv. Bartlett and Packham's Triumph to improve hand thinning. *Acta. Hortic.* 475. 283-293.

- [2] Gimenez, G. N. (2012). Determinación de los patrones de crecimiento y abscisión de frutos para ajustar el momento oportuno de raleo en perales cv. 'Williams'. Tesis de Maestría en Fruticultura de Clima Templado-Frío, Universidad Nacional del Comahue.
- [3] Lindstrom, M. L. & Bates, D. M, (1990). Nonlinear mixed effects models for repeated measures data. *Biometrics*, 46(3):673-687.
- [4] Ratkowski, D.A. (1983). *Nonlinear Regression Modeling: a Unified Practical Approach*, Marcel Dekker Inc., NY,USA.
- [5] Tassile, V. (2020). Aplicación de modelos mixtos no lineales en el crecimiento de frutos de peral (*Pyrus communis*) cv. 'William's'. Tesis de Maestría, Universidad Nacional de Córdoba.

## **APLICACIÓN DEL MODELO DE REGRESIÓN LOGÍSTICA PARA EL ESTUDIO DE LA DESERCIÓN ESTUDIANTIL EN LOS INGRESANTES 2019 DE LA FACULTAD DE CIENCIAS ECONÓMICAS, UNIVERSIDAD NACIONAL DE ENTRE RÍOS**

D'IORIO STEFANÍA<sup>1</sup>, AVILA OLGA B.<sup>1</sup>

<sup>1</sup> *Facultad de Ciencias Económicas, Universidad Nacional de Entre Ríos (UNER), Paraná, Entre Ríos Argentina*  
*stefaniadiorio@fceco.uner.edu.ar*

### **RESUMEN**

La deserción estudiantil es uno de los problemas más complejos y frecuentes que enfrentan las Universidades de nuestro país, Argentina, surgiendo la necesidad de investigar los factores que influyen en el abandono de los ingresantes a las carreras universitarias. En consecuencia, el presente trabajo tuvo como objetivo identificar los factores que influyen en la deserción de los ingresantes 2019 a las carreras de grado y pregrado de la Facultad de Ciencias Económicas de la Universidad Nacional de Entre Ríos. Para ello se utilizó un modelo de regresión logística donde se considera la influencia de diferentes variables en la reinscripción de los estudiantes al segundo año de la carrera, en este caso al año académico 2020, considerando esta variable como posible indicador de deserción.

El modelo logístico resulta apropiado ya que la variable dependiente es dicotómica, y permite pronosticar qué sucederá cuando existan dos posibilidades: se reinscribió o no al siguiente año académico. Como posibles variables predictoras se consideraron Sexo, Edad al finalizar el período, Cantidad de inscripciones a cursadas en el año, y Cantidad de materias aprobadas, corriéndose un modelo para las carreras de grado y otro para las carreras de pregrado. Se trabajó con el total de ingresantes del año académico 2019, que fueron 556 para las carreras de grado y 628 para las carreras de pregrado. Los dos modelos propuestos presentaron un buen ajuste, y ambos proporcionaron una buena clasificación para los casos estudiados, 80% el de grado y 83,8% el de pregrado, permitiendo detectar diferentes variables estadísticamente significativas en el estudio de la reinscripción de los estudiantes en el año 2020.

**Palabras clave:** *regresión logística, deserción estudiantil, ingresantes universitarios.*

### **INTRODUCCIÓN**

El problema de la deserción estudiantil es uno de los problemas más complejos y frecuentes que enfrentan las Universidades de nuestro país, Argentina, lo que resulta en la necesidad de realizar investigaciones para identificar los factores que influyen en el abandono de los ingresantes a las carreras universitarias. Los resultados de estas investigaciones servirían para identificar y atender las causas que intervienen en el abandono, y poder ejecutar políticas universitarias tendientes a la retención de los ingresantes.

La Facultad de Ciencias Económicas de la Universidad Nacional de Entre Ríos, que históricamente ofreció únicamente la carrera de Contador Público, hace algunos años ha ampliado su oferta universitaria, agregando las carreras de Licenciatura en Economía y Licenciatura en Gestión de las Organizaciones, e incorporando recientemente carreras de pregrado, como la Tecnicatura Universitaria en Gestión de los Recursos Humanos, la

Tecnicatura Universitaria en Gestión y Administración Pública, y la Tecnicatura Universitaria Administrativo Contable. Estas carreras son independientes entre sí.

Ante esta expansión de la oferta académica de la Facultad, surgió como necesidad poder contar con información sobre los estudiantes y su rendimiento académico, creándose el Programa de Producción de Estadísticas Universitarias, con el objetivo de proporcionar información oportuna para la toma de decisiones de planificación y gestión de políticas educativas de esta unidad académica. Dentro de este programa ha surgido como necesidad el análisis específico del problema de la deserción estudiantil.

La deserción puede explicarse como el resultado de distintos factores que afectan al estudiante. En este trabajo, y siguiendo la clasificación de Tinto (1989) sobre los períodos críticos en la trayectoria estudiantil, se analiza la deserción en el tercer período, que se da entre la última fase del primer año de estudios y antes del comienzo del segundo. Consecuentemente, el presente trabajo tuvo como objetivo identificar los factores que influyen en la deserción de los ingresantes 2019 a las carreras de grado y pregrado de la Facultad de Ciencias Económicas de la Universidad Nacional de Entre Ríos. Para ello se utilizó un modelo de regresión logística donde se considera la influencia de diferentes variables en la reinscripción de los estudiantes al segundo año, en este caso, la reinscripción al año académico 2020.

## **MATERIALES Y MÉTODOS**

Para estudiar el mejor modelo que predice la probabilidad que un estudiante abandone la universidad se utiliza en este trabajo el modelo de regresión logística.

Se considera como variable dependiente: la reinscripción al año académico 2020, considerando esta variable como posible indicador de deserción. Y como variables independientes: Sexo, Edad al finalizar el período, Cantidad de inscripciones a cursadas en el año y Cantidad de materias aprobadas, constatándose previamente la no multicolinealidad entre las mismas. Se corrieron dos modelos: uno para los estudiantes de las carreras de grado y otro para los estudiantes de las carreras de pregrado.

Para realizar las predicciones, este modelo se considera apropiado ya que tiene en cuenta una variable dependiente dicotómica (se reinscribe o no al año académico 2020) y se quiere estudiar el efecto que diferentes variables independientes tienen sobre ella. También llamado modelo de respuesta cualitativa, tiene utilidad para pronosticar qué sucederá cuando existan dos posibilidades, en nuestro caso, se reinscribió o no al siguiente año académico. Este modelo permite, por un lado, estimar la probabilidad que un estudiante tiene de abandonar la universidad dada la presencia de determinadas variables independientes y, por otro, evaluar la influencia que cada variable independiente tiene en la respuesta en forma de OR (ODD RATIO). Una OR mayor que uno indica aumento en la probabilidad del evento, y una OR menor que uno implica una disminución.

Para la estimación de los coeficientes en el modelo se utilizó el método de máxima verosimilitud, mientras que para ver la de bondad de ajuste del modelo se consideró la prueba de Hosmer y Lemeshow y la prueba que considera  $-2\log$  de la verosimilitud. Se trabajó con el total de ingresantes a las carreras mencionadas del año académico 2019, que fueron 556 para las carreras de grado y 628 para las carreras de pregrado. El porcentaje de reinscriptos al año académico siguiente fue de 60,6% para los estudiantes de grado, y 54,9% para los estudiantes de pregrado.

## RESULTADOS Y DISCUSIÓN

En el modelo logístico para las carreras de grado se obtuvieron como variables significativas estadísticamente Materias aprobadas y Cantidad de inscripciones a cursadas en el año. Ambas variables tienen una OR mayor a uno, lo que indica un aumento en la probabilidad de deserción estudiantil. Los resultados de las variables del modelo se presentan a continuación:

	B	Error estándar	Sig.	Exp(B)	95% C.I. para EXP(B)	
					Inferior	Superior
Edad	,010	,023	,661	1,010	,965	1,057
Materias aprobadas	,258	,060	,000	1,294	1,151	1,454
Sexo	,005	,229	,982	1,005	,642	1,575
Cantidad de inscripciones a cursadas en el año	,424	,047	,000	1,528	1,394	1,676
Constante	-2,313	,669	,001	,099		

Con el modelo para las carreras de grado el porcentaje de clasificación de los casos de manera correcta es de 80%.

En el modelo logístico para las carreras de pregrado, todas las variables independientes resultaron significativas estadísticamente, lográndose con este modelo un 83,8% de clasificación correcta de los casos. Las variables Materias aprobadas, Sexo y Cantidad de inscripciones a cursadas en el año presentan una OR mayor a 1, indicando un aumento en la probabilidad de deserción estudiantil, mientras que lo contrario sucede con la variable Edad. Los resultados de las variables del modelo se presentan a continuación:

	B	Error estándar	Sig.	Exp(B)	95% C.I. para EXP(B)	
					Inferior	Superior
Edad	-,027	,013	,044	0,973	,948	0,999
Materias aprobadas	,696	,070	,000	2,006	1,748	2,302
Sexo	,604	,248	,015	1,830	1,126	2,975
Cantidad de inscripciones a cursadas en el año	,173	,042	,000	1,188	1,095	1,290
Constante	-1,694	,496	,001	,184		

Los test de bondad de ajuste de ambos modelos indican un buen ajuste. Si bien periódicamente se analiza el porcentaje de deserción de los ingresantes a través de los Anuarios Estadísticos elaborados en el marco del Programa de Estadísticas Universitarias, a los fines de diagnosticar y tomar decisiones en busca de la atenuación de este problema es que se ha elegido aplicar esta técnica de regresión logística, que logra identificar un conjunto de variables que caracterizan la deserción estudiantil de los ingresantes del año 2019 de la Facultad de Ciencias Económicas de la Universidad Nacional de Entre Ríos.

Ambos modelos de regresión logística aplicados han proporcionado una buena clasificación para los casos estudiados. Asimismo, esta técnica permitió detectar diferentes variables estadísticamente significativas en el estudio de la reinscripción de los estudiantes en el año 2020, utilizada como posible variable indicadora de deserción de los ingresantes 2019.

Se va a continuar con el estudio de la deserción estudiantil de los ingresantes de la Facultad, tratando de incorporar variables socioeconómicas que contribuyan a detectar otras posibles causas de abandono de estudios. Todo ello para aportar a la planificación, gestión y ejecución de políticas universitarias tendientes a la retención de los ingresantes.



## **REFERENCIAS**

- [1] Hosmer, D. Lemeshow, S. (2000). Applied Logistic Regression. John Wiley and Sons Inc.
- [3] Tinto, V (1989). Definir la deserción: una cuestión de perspectiva. *Revista de Educación Superior* N° 71, ANUIES, México.
- [3] Díaz Peralta, C. (2008). Modelo conceptual para la deserción estudiantil universitaria chilena. *Estudios pedagógicos (Valdivia)*, 34(2), 65-86.

## IMPACTO DEL USO DE DISTINTOS ESTÁNDARES DE REFERENCIA PARA EL CÁLCULO DE LA PREVALENCIA DE RETRASO EN CRECIMIENTO

DRAMIS AGUSTÍN NICOLÁS<sup>1</sup>, ALTSZYLER EDGAR<sup>2</sup>, TURJANSKI PABLO GUILLERMO<sup>3</sup>, CUETO GERARDO RUBÉN<sup>4</sup>, FERNANDEZ MARÍA SOLEDAD<sup>5</sup>, NUÑEZ PABLO<sup>6</sup>, PEREZ ADRIANA ALICIA<sup>7</sup>

<sup>1,2,3,4,5,6,7</sup>Grupo de Bioestadística Aplicada, (DEGE/DC/ICC/IC)-FCEN- UBA/CONICET

### RESUMEN

El retraso en el crecimiento, una forma de desnutrición infantil, es una problemática grave asociada a consecuencias negativas en la salud, la educación y la condición socioeconómica en la vida adulta. Su diagnóstico se realiza mediante el cálculo de puntajes estandarizados o  $z$  para un individuo respecto a una curva de referencia de talla para la edad y el sexo. Existen diversas fuentes, nacionales e internacionales, de curvas de crecimiento a distintas edades. Dado que los patrones de crecimiento pueden diferir entre distintas regiones, el uso de ciertas curvas podría generar sesgos, sobre o subestimando la proporción de individuos con esta forma de desnutrición. Para evaluar esto, se analizó el impacto de la elección de la curva de referencia sobre la prevalencia de retraso en el crecimiento en establecimientos de salud pertenecientes al programa SUMAR en todo el país entre enero y agosto de 2019. Se observó una sobreestimación de la prevalencia de retraso en el crecimiento al utilizar las curvas de referencia internacionales respecto a las nacionales para individuos mayores a 6 años, profundizada para mayores de 10. No se observaron diferencias relevantes al comparar las curvas nacionales respecto a los estándares multicéntricos de la OMS. Los resultados concuerdan con la hipótesis de dinámicas de crecimiento distintas en distintas regiones, y apoyan las recomendaciones de la Sociedad Argentina de Pediatría.

**Palabras clave:** *nutrición infantil, regresión tipo II, estandarización.*

### INTRODUCCIÓN

La desnutrición infantil es una grave problemática actual. Sus distintas formas se asocian a mayor mortalidad y carga de enfermedades en la niñez, así como a una pérdida de capital económico en la vida adulta<sup>1</sup>. En particular, el retraso en el crecimiento, que refleja una baja talla para la edad, es considerado un buen indicador de desnutrición crónica<sup>2</sup> y se encuentra asociado a perjuicios en la salud, educación y desarrollo económico a lo largo de la vida<sup>3</sup>.

El diagnóstico de retraso en el crecimiento se realiza calculando el puntaje estandarizado o  $z$  de un individuo respecto a una curva de referencia de talla esperada para la edad y el sexo. Se considera desnutrido a un individuo con un puntaje  $z$  menor a  $-2$ . Existen diferentes fuentes de curvas de referencia de crecimiento en talla: a) Organización mundial de la Salud (OMS) multicéntrica: de 0 a 5 años, generada a partir de una muestra de niños y niñas de distintos países en buenas condiciones de salud y alimentación<sup>4</sup>; b) National Centers for Health Statistics/Centers for Disease Control and Prevention (NHCS/CDC) internacional: entre 5 y 18 años, construida a partir de una muestra aleatoria de niños, niñas y adolescentes de Estados Unidos<sup>5</sup>; c) Sociedad Argentina de Pediatría (SAP) nacional: entre 0 y 19 años, generadas a partir de muestras de niños y niñas de Argentina<sup>6</sup>. La utilización de la referencia OMS

multicéntrica para menores de 5 años está ampliamente aceptada, dado que no se observan grandes diferencias en el crecimiento entre las poblaciones de distintas regiones del mundo<sup>6</sup>. Sin embargo, existen evidencias de que alrededor de la adolescencia las poblaciones pueden presentar distintos patrones de crecimiento lineal, limitando la aplicabilidad de las referencias internacionales<sup>7</sup>.

Resulta de importancia determinar el impacto que tienen estas diferencias en la dinámica de crecimiento sobre el diagnóstico de la desnutrición en Argentina. Para ello, el objetivo de este trabajo fue comparar las prevalencias de desnutrición en establecimientos públicos de salud de Argentina utilizando referencias nacionales e internacionales. Dicho análisis permitiría minimizar sesgos a la hora de evaluar la situación nutricional de los niños, niñas y adolescentes argentinos.

## **MATERIALES Y MÉTODOS**

La población de estudio consistió en niños, niñas y adolescentes de 0 a 19 años, sin obra social ni seguro médico, beneficiarios del Programa SUMAR (Ministerio de Salud de la Nación), que se efectuaron controles entre enero y agosto de 2019 en algún establecimiento público de salud del país. Estos controles incluyen datos antropométricos de talla, además de sexo, fecha de nacimiento y fecha del control. Previo al análisis, se realizó un proceso de limpieza de la base, depurando observaciones inconsistentes o con valores implausibles. Este proceso fue realizado procurando minimizar la pérdida de datos.

Para cada registro de talla se calculó el puntaje z según distintas curvas de referencia (OMS multicéntrica, NHCS/CDC internacional y SAP nacional). A partir de los datos de puntaje z, se calculó la prevalencia de retraso en el crecimiento para cada establecimiento de salud. En caso de que un individuo presentase más de un registro en el período estudiado, se consideró el último. Dado que los establecimientos con baja cantidad de individuos tienden a presentar proporciones más extremas, estas fueron graficadas en función del n del establecimiento, y en función de esto se decidió incluir en el análisis solamente a los establecimientos con una cantidad de individuos mayor a 250.

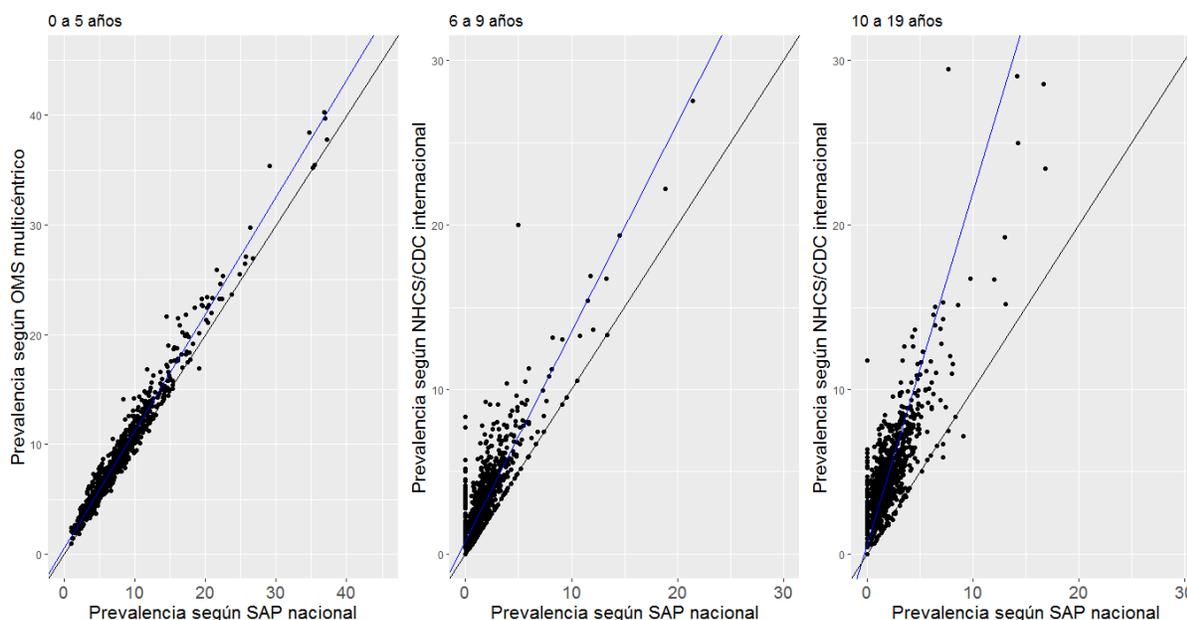
La comparación entre prevalencias calculadas según las distintas referencias se efectuó estratificando rango etario. Se efectuaron análisis gráficos, determinando la pendiente observada mediante regresión de tipo II. Todos los análisis fueron realizados con el software R8 versión 3.6.0.

## **RESULTADOS Y DISCUSIÓN**

Se analizaron datos de prevalencias de retraso en el crecimiento en 6445 establecimientos de salud. Al realizar el filtrado por tamaño mínimo, se retuvieron entre 1131 y 1433 establecimientos según el rango de edades. Se observó una tendencia a registrar una mayor prevalencia de retraso en el crecimiento al utilizar como referencia la fuente NHCS/CDC internacional respecto a utilizar la fuente SAP nacional. Esta tendencia resultó más marcada a edades más avanzadas (pendientes: 1,27 para 6 a 9 años; 2,15 para 10 a 19). No se observó una diferencia relevante al comparar entre las fuentes OMS multicéntrica y SAP nacional para menores de 5 años (pendiente: 1,07) (Figura 1).

Los resultados apoyan la hipótesis del conflicto de aplicabilidad de estándares internacionales a poblaciones nacionales, debido a la distinta dinámica de crecimiento<sup>7</sup>. Utilizar las curvas de NHCS/CDC generaría una sobreestimación de entre un 27 a un 115% de la prevalencia de retraso en el crecimiento en poblaciones argentinas de mayores de 6 años. Para los menores de 5 años no se observaron diferencias relevantes entre los estándares de crecimiento de la

OMS y las referencias nacionales. Esto puede deberse a menores diferencias en la dinámica de crecimiento a edades tempranas, o a un menor efecto de diferencias regionales debido a la naturaleza multicéntrica del estudio de la OMS. Los resultados obtenidos apoyan las recomendaciones de la Sociedad Argentina de Pediatría, según las cuales se debe utilizar la referencia SAP nacional para mayores de 6 años.



**Figura 1.** Comparación de prevalencias según fuentes de referencia. Las líneas azules representan regresión de tipo II; las líneas negras representan una pendiente igual a 1.

## REFERENCIAS

- [1] Black, R. E., Allen, L. H., Bhutta, Z. A., Caulfield, L. E., De Onis, M., Ezzati, M., Mathers, C., Rivera, J. & Maternal and Child Undernutrition Study Group. (2008). Maternal and child undernutrition: global and regional exposures and health consequences. *The lancet*, 371(9608), 243-260.
- [2] World Health Organization. (1995). *Physical status: The use of and interpretation of anthropometry, Report of a WHO Expert Committee*. World Health Organization.
- [3] McGovern, M. E., Krishna, A., Aguayo, V. M., & Subramanian, S. V. (2017). A review of the evidence linking child stunting to economic outcomes. *International journal of epidemiology*, 46(4), 1171-1191.
- [4] World Health Organization. (2006). *WHO child growth standards: length/height-for-age, weight-for-age, weight-for-length, weight-for-height and body mass index-for-age: methods and development*. World Health Organization.
- [5] Kuczmarski, R. J. (2000). *CDC growth charts: United States* (No. 314). US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics.
- [6] Sociedad Argentina de Pediatría (2013). *Guía para la evaluación del crecimiento físico*. Buenos Aires: Sociedad Argentina de Pediatría. Argentina



- [7] Orden, A. B., & Apezteguía, M. C. (2016). Weight and height centiles of Argentinian children and adolescents: a comparison with WHO and national growth references. *Annals of human biology*, 43(1), 9-17.
- [8] R Team Core (2014). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org>

## EXPLORACIÓN DE EFECTOS AMBIENTALES SOBRE EL TAMAÑO DE CAMADA EN TRES LÍNEAS DE CONEJOS

FERNÁNDEZ E.N.<sup>1</sup>, MARTINEZ R.D.<sup>1</sup>, ABBIATI N.N.<sup>1</sup>

<sup>1</sup>Facultad de Ciencias Agrarias, IIPAS, Universidad Nacional de Lomas de Zamora  
ednfer@yahoo.com

### RESUMEN

El desarrollo de una línea materna de conejos para carne comienza con el establecimiento de una generación inicial, cuyos animales se asumen no emparentados. En la evaluación de conejos pueden considerarse los efectos ambientales fijos: estaciones, orden de parto y el estado fisiológico de la hembra al momento de monta efectiva, entre otros. Este trabajo tiene como objetivo explorar modelos con efectos ambientales sobre la variable número de destetados, en la generación inicial de 3 líneas de conejos para carne (LA, LV y LH). Se propusieron distintos modelos, explorando la correlación entre mediciones de una misma coneja y las posibles heterogeneidades en las varianzas de los errores asociadas a los niveles de los distintos efectos fijos, empleando algoritmos de los modelos mixtos. Manteniendo los efectos fijos, el mejor modelo para la matriz residual consideró: Simetría Compuesta y heterogeneidad de varianza para el estado fisiológico en LA; en LV, heterogeneidad de varianzas para las estaciones y en LH una estructura Autorregresiva de primer orden y heterogeneidad de varianza para los estados fisiológicos. Se detectaron diferencias en las medias del estado fisiológico para LA y LH, en las estaciones para LV y en orden de parto para LH.

**Palabras clave:** *modelos mixtos, efectos ambientales, conejos, prolificidad.*

### INTRODUCCIÓN

El desarrollo de una línea materna de conejos para carne comienza con el establecimiento de una generación inicial, elegida por algún criterio, cuyos animales se asumen no emparentados (población base). Sobre esta, se desarrolla un proceso selectivo que requiere poder discriminar los factores genéticos y ambientales que afectan al carácter de interés. En particular, en la evaluación de conejos pueden considerarse los efectos ambientales fijos: años, estaciones (E), orden de parto (OP) y el estado fisiológico de la hembra al momento de monta efectiva (EF), entre otros. En las líneas maternas de conejos los objetivos de selección se centran en los caracteres reproductivos, siendo uno de los más importantes el número de gazapos destetados (ND). Este rasgo, puede medirse sucesivamente en los diferentes partos de una misma coneja, pudiendo estar afectados por distintos efectos ambientales. Los errores correspondientes a los niveles de estos efectos pueden además diferir en cuanto a su variabilidad. Los modelos lineales mixtos no sólo permiten contemplar efectos fijos y aleatorios, sino que posibilitan modelar la matriz residual (R) tanto en su diagonal, incorporando heterogeneidad de varianzas, como en sus elementos no diagonales, incluyendo la correlación entre mediciones sucesivas de un mismo individuo (Wang y Goonewardene, 2004). Este trabajo tiene como objetivo explorar modelos con efectos ambientales sobre ND, en la población base de 3 líneas de conejos para carne.

## MATERIALES Y MÉTODOS

Se dispuso de una base de registros para ND provenientes de 3 líneas (LA, LV y LH) pertenecientes a la Universidad Politécnica de Valencia, España. El número de registros y cantidad de conejas fue de: 218 (92) para LA, 144 (73) para LV y 459 (136) para LH. Se consideraron tres efectos, E (V: Verano, O: Otoño, I: Invierno y P: Primavera), OP (4 niveles para LV y 6 niveles LA y LH) y el EF (nulíparas (1), primíparas lactantes (2), multíparas lactantes (3), primíparas no lactantes (4) y multíparas no lactantes (5)). En LV no se dispuso de datos para P. Se propusieron distintos modelos candidatos para ND, manteniendo los efectos fijos y explorando la correlación entre mediciones de una misma coneja y las posibles heterogeneidades de los errores. Se emplearon algoritmos de los modelos mixtos y se usó el Criterio de Información de Akaike (AIC) para la comparación de modelos. Se trabajó con un nivel de significación  $\alpha$  del 0.05 y la prueba DGC para la comparación de medias. Se usó la librería nlme del lenguaje R, implementada en InfoStat (Di Rienzo *et al.*, 2017).

## RESULTADOS Y DISCUSIÓN

Con relación a E, el mayor porcentaje de registros se presentó en O para LA (50%) y LH (41%) y en V para LV (50%). En cuanto al EF, la mayor frecuencia de registros correspondió a las Nulíparas en LA (30%) y LV (49%) y Multíparas Lactantes para LH (41%). Respecto a la cantidad máxima de partos por coneja, en LA lo más frecuente fue de 2 (34%), en LV de 1 (44%) y en LH de 5 (20%), siendo esta línea la más homogénea en este aspecto.

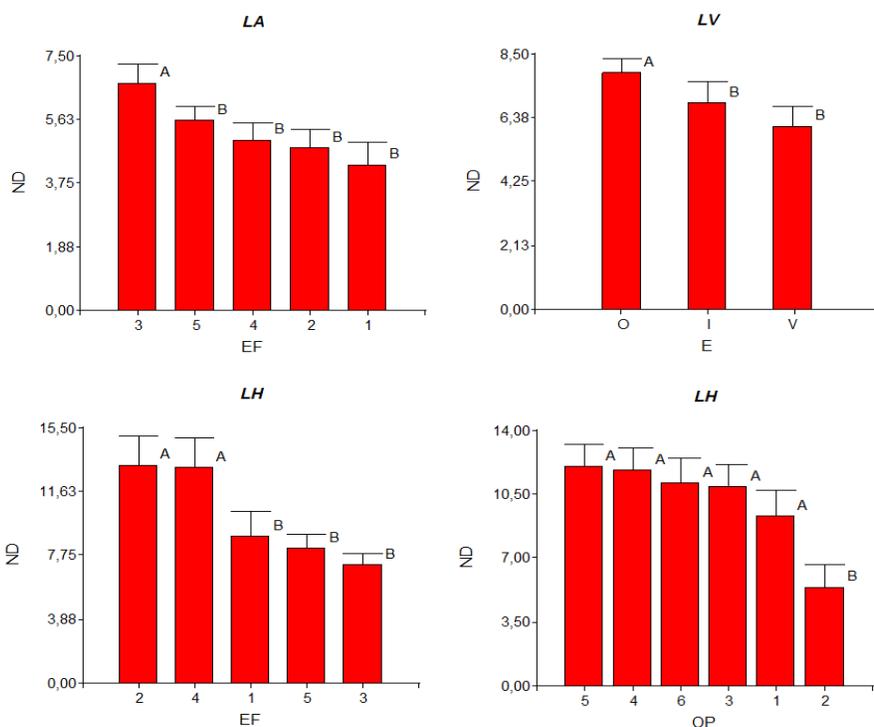
El mejor modelo para LA en relación con la matriz R fue Simetría Compuesta (SC) para la correlación entre partos sucesivos de una misma coneja y heterogeneidad de varianza para los residuos de las observaciones agrupadas según EF. En LV, resultó un modelo sin correlación entre partos y heterogeneidad de varianzas para los residuos agrupados según E. Para LH el modelo elegido contempló una estructura Autorregresiva de primer orden (AR1) para los partos sucesivos y heterogeneidad de varianza para los residuos de las observaciones agrupadas por EF. El Cuadro 1 muestra los valores p asociados a los efectos fijos con los AIC correspondientes y el Cuadro 2 presenta las estimaciones referidas a R.

Cuadro1: Valores p asociados a los efectos fijos y AIC en cada línea

Línea	E	EF	OP	AIC
LA	0,4011	0,0409	0,7039	1007,12
LV	0,0111	0,8100	0,7496	635,54
LH	0,1637	0,0183	0,0433	2237,09

Cuadro 2: Correlaciones y varianzas estimadas en cada línea

Línea	Varianza del error		Correlación
	Agrupamiento: EF		Simetría Compuesta
LA	1	4,66	0,47
	2	9,14	
	3	4,57	
	4	7,17	
	5	7,88	
LV	Agrupamiento: E		Independiente
	1	7,95	0
	2	2,15	
3	5,09		
LH	Agrupamiento: EF		Autorregresivo ord. 1
	1	5,81	0,15
	2	7,95	
	3	9,08	
	4	9,82	
5	5,02		



Prueba DGC. Medias con las mismas letras no difieren.

Gráfico 1: Medias y errores estándares de los efectos fijos significativos

En el Gráfico 1 se presentan los resultados de la prueba DGC para los efectos fijos significativos en cada línea. En LA, el EF múltiparas lactantes presenta un promedio mayor que el resto. En LV la estación Otoño se diferenció de las restantes. En LH las primíparas mostraron medias mayores y el OP 2, resultó significativamente menor.

La exploración de los efectos fijos en la etapa fundacional de las líneas evidenció que éstas se comportan de manera diferente. El EF múltiparas lactantes se comportó de manera opuesta en LA y LH. Además, las estaciones sólo mostraron diferencias en LV y el OP en LH. En cuanto a las mediciones repetidas, LV no necesitó incorporar una estructura de correlación entre OP, probablemente debido a la menor cantidad de información disponible, contrariamente a lo sucedido con LA y LH. Si comparamos estas estimaciones con las obtenidas en otros análisis que incluyeron 37, 34 y 15 generaciones para LA, LV y LH y los animales como efectos aleatorios (considerando la genealogía completa) podemos observar que, respecto a la correlación entre medidas repetidas, las estimaciones son coincidentes para LH (0,15), superior en LA (0,47 vs 0,16) y menor en LV (0 vs 0,16), (Fernández, 2016).

## **REFERENCIAS**

- [1] Di Rienzo, J. A., Macchiavelli, E., y Casanoves, F. (2017). *Modelos lineales mixtos: aplicaciones en InfoStat*. Actualización 2017, 1a Edición especial, Argentina.
- [2] Fernández, E. N. (2016). Estimación de efectos genéticos aditivos y no aditivos en líneas maternas de conejos (Doctoral dissertation).
- [3] Wang, Z. y Goonewardene, L. A. (2004). "The use of MIXED models in the analysis of animal experiments with repeated measures data". *Canadian Journal of Animal Science*, 84(1): 1-11.

## MODELACIÓN DE C Y N PROVENIENTE DE BIOMASA DE RAÍCES EN FUNCIÓN DE LA PROFUNDIDAD DEL SUELO

GILI ADRIANA A.<sup>1</sup>, BELMONTE VALERIA<sup>1</sup>, GARCÍA MARÍA V.<sup>1</sup>, FRASIER ILEANA<sup>2,3</sup>, NOELLEMEYER ELKE J.<sup>4</sup>

<sup>1</sup>Facultad de Agronomía, Cátedra de Estadística, UNLPam

<sup>2</sup>Estación Experimental Guillermo Covas, INTA Anguil

<sup>3</sup>Consejo Nacional de Investigaciones Científicas y Técnicas

<sup>4</sup>Facultad de Agronomía, Cátedra de Edafología, UNLPam  
agili@agro.unlpam.edu.ar

### RESUMEN

Las raíces representan un aporte importante de residuos al suelo, contribuyendo de manera eficiente al secuestro de nutrientes y, por lo tanto, a mejorar la calidad del mismo. Los contenidos de C y N de la biomasa de raíces (C-BRT y N-BRT) disminuyen con la profundidad, de forma no lineal. En el presente trabajo se propone modelar la relación entre los contenidos de C-BRT y N-BRT y la profundidad del suelo, bajo diferentes usos y clases texturales en suelos de la región semiárida pampeana. Además, comparar el ajuste de un modelo no lineal mixto con la aproximación basada en coeficientes fijos. Para modelar la relación se consideró un modelo potencial. Se evaluaron modelos con coeficientes fijos, con coeficientes fijos y aleatorios e incluyendo variables auxiliares para comparar los distintos usos del suelo. Además, se modeló la presencia de heterocedasticidad en los modelos ajustados. Se usó el programa R y el paquete NLME. Los modelos de coeficientes fijos fueron los mejores desde el punto de vista estadístico y permitieron obtener las estimaciones de los parámetros que definen la relación estudiada en forma conjunta para cada uno de los usos del suelo, dentro de cada clase textural.

**Palabras clave:** *modelos no lineales mixtos, nlme, gnls.*

### INTRODUCCIÓN

El desarrollo de estrategias de manejo del uso del suelo para mejorar el almacenamiento de carbono (C) y nitrógeno (N) en el suelo, se ha convertido en un desafío actual para mitigar el cambio climático y recuperar los suelos degradados (Poirier *et al.*, 2018; Sokol *et al.*, 2018). Las raíces representan un aporte importante de residuos al suelo contribuyendo de manera eficiente al secuestro de C y N y, por lo tanto, a mejorar la calidad del suelo. Los contenidos de C y N de la biomasa de raíces (C-BRT y N-BRT) disminuyen con la profundidad, sin embargo, dicho comportamiento no resulta ser lineal. Los modelos no lineales permiten modelar datos con mucha flexibilidad y menos parámetros (más parsimoniosos). Los modelos no lineales mixtos, permiten que los efectos fijos y los aleatorios tengan una relación no lineal con la variable de respuesta. Estos modelos pueden tomar varias formas, pero los más comunes involucran una distribución condicional para la variable en estudio dados los efectos aleatorios (Wolfinger, 1998).

En el presente trabajo se propone modelar la relación entre los contenidos de C-BRT y N-BRT y la profundidad considerada en el suelo bajo diferentes usos y clases texturales. Además, comparar la aproximación sujeto específica (modelo no lineal mixto) con la aproximación

basada en coeficientes fijos (modelo no lineal) con el fin de identificar cual fue la mejor opción para este caso en estudio.

## MATERIALES Y MÉTODOS

**Datos:** El área de estudio corresponde a la región semiárida central de Argentina. Dentro del área de estudio se evaluaron 21 sitios con diferentes usos del suelo y texturas. Las texturas consideradas son: franco, franco arenosos y arenosos francos. En cuanto a los usos, se seleccionaron: sitios agrícolas (A), sitios con vegetación natural (NG): pastizales y monte de caldén, y sitios con pasturas permanente (PP): pasturas de pasto Llorón y/o Panicum, Los contenidos de C y N de la biomasa de raíces fueron determinados por combustión seca (LECO-TrueSpec®).

**Modelación estadística:** La relación entre la variable y (C-BRT o N-BRT) y la profundidad se modeló a partir de un modelo potencial (1).

$$y_{ij} = \frac{\beta_1}{\text{prof}\beta_2} + \varepsilon_{ij} \quad (1)$$

$\beta_1$  controla la tasa de aumento o disminución de la curva.  $\beta_2$  controla la forma de la curva ( $\beta_1 > 0$ ), cuando es menor a cero, determina una curva cóncava hacia arriba y decreciente. Luego considerando como efectos aleatorios los distintos puntos de muestreo se va a considerar el siguiente modelo no lineal de efectos mixtos:

$$y_{ij} = \frac{\phi_1}{\text{prof}\phi_2} + \varepsilon_{ij} \quad (2)$$

donde  $\phi_{1i} = \beta_1 + b_{1i}$ ;  $\phi_2 = \beta_2 + b_{2i}$ ;  $b_i \sim N(0, \Psi)$ ,  $\varepsilon_{ij} = N(0, \sigma^2)$

Se agregó una covariable para estudiar los distintos usos del suelo, quedando el modelo final en función del presentado en (2), cambiando la formulación de  $\phi_{1i}$  y  $\phi_2$  como:

$$\begin{aligned} \phi_{1i} &= \beta_1 + \gamma_1 D_{1i} + \gamma_2 D_{2i} + b_{1i} \\ \phi_2 &= \beta_2 + \rho_1 D_{1i} + \rho_2 D_{2i} + b_{2i} \end{aligned} \quad (3)$$

donde  $\beta_1$  y  $\beta_2$  representan el valor promedio del parámetro para la categoría de referencia,  $\gamma_1$  y  $\gamma_2$  representan la desviación de la estimación del valor del parámetro con respecto al valor de su parámetro de referencia.  $D_{1i}$  y  $D_{2i}$  son las variables auxiliares usadas para incluir en el modelo la variable tipo de uso del suelo.

Ante la falta del cumplimiento de los supuestos planteados sobre el error se evaluaron también modelos con errores heterocedásticos y correlacionados dentro de cada punto de muestreo. El ajuste de modelos se realizó a través de máxima verosimilitud usando las funciones gnls y nlme del paquete NLME (Pinheiro *et al.*, 2020) del programa R (R Core Team, 2018). Los pasos de la modelación fueron: en primer lugar, se comenzó con el modelo saturado de efectos fijos y parámetros de las distribuciones de efectos aleatorios, luego se eligieron los efectos aleatorios importantes para simplificar la estructura y finalmente a partir de la estructura de efectos aleatorios seleccionados, se consideraron los efectos fijos importantes.

Los diferentes modelos ajustados se compararon a través de pruebas de cocientes de verosimilitud (LRT) y de los criterios AIC y BIC. Los resultados presentados corresponden a los modelos elegidos como los mejores para caracterizar los datos.

## RESULTADOS Y DISCUSIÓN

Los modelos de efectos fijos fueron los que mejor explicaron las relaciones entre los contenidos de C-BRT y N-BRT con la profundidad. La inclusión de coeficientes aleatorios no mejoró los modelos y dado que lo que se busca es interpretar el efecto general de los usos del suelo sobre la relación estudiada, poder contar con la esperanza marginal fue lo más adecuado en este caso (es decir, el efecto para el promedio de la población). Se observó una marcada

heterocedasticidad y la misma fue modelada a través de la función potencia (VarPower). La tasa de disminución de C-BRT con la profundidad en los suelos con textura franco y arenoso franco fue marcadamente superior en el uso A (Tabla 1).

En el suelo franco con el uso A se observa el mayor valor estimado de  $\beta_2$ , lo que indica que el cambio con la profundidad es más abrupto (Figura 1), esto es explicado por la característica de los cultivos agrícolas que tienden a concentrar las raíces mayoritariamente en el estrato superior de suelo. Por el contrario, para el uso NG es mucho más gradual, dado que hay una mayor diversidad de especies con diferentes características de desarrollo y exploración radical del suelo. Con respecto a N-BRT (Tabla 2) presenta para las tres texturas el menor valor estimado de  $\beta_2$ , lo que indica una gradualidad mayor al variar la profundidad que los otros dos usos.

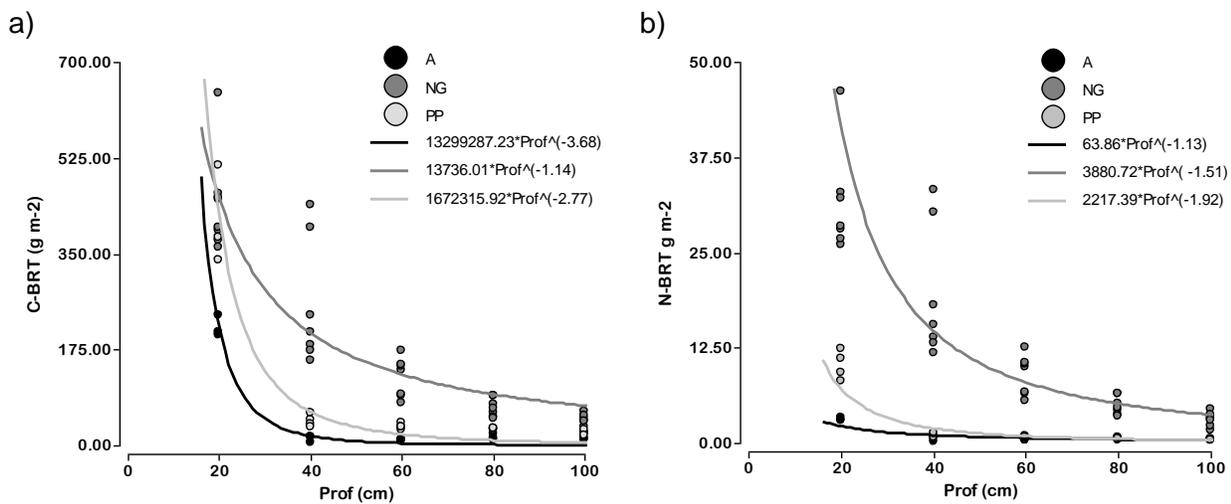


Figura 1. Comparación de modelos ajustados para a) C-BRT y b) N-BRT en relación con la profundidad para cada uso considerado. A: agrícola, NG: vegetación natural y PP: pastura perenne.

Tabla 1. Estimación de los coeficientes que explican la relación entre C-BRT y la profundidad teniendo en cuenta los distintos usos, para cada textura por separado.

	Franco		Franco arenoso		Arenoso franco	
	b1	b2	b1	b2	b1	b2
A	13299287.23 c	-3.68 c	91424.61 b	-2.07 b	2530483.63 a	-3.06 b
NG	13736.01 a	-1.14 a	15969.66 a	-1.24 a	48610.84 b	-1.68 a
PP	1672315.92 b	-2.77 b	1160436.75 c	-2.66 c	89641.08 c	-2.02 b

Letras diferentes indican diferencias significativas entre usos para cada parámetro (p-valor<0.05).

Tabla 2. Estimación de los coeficientes que explican la relación entre N-BRT y la profundidad teniendo en cuenta los distintos usos, para cada textura por separado.

	Franco		Franco arenoso		Arenoso franco	
	a	b	a	b	a	b
A	63.86 a	-1.13 a	118.14 a	-1.28 a	4246.08 ab	-2.34 b



NG	3880.72 b	-1.51 a	2314.17 b	-1.50 a	8137.66 b	-1.93 a
PP	2217.39 b	-1.91 b	48286.10 c	-2.85 b	1744.50 a	-1.95 a

Letras diferentes indican diferencias significativas entre usos para cada parámetro (p-valor<0.05).

## REFERENCIAS

- [1] Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., R Core Team, 2020. {nlme}: Linear and Nonlinear Mixed Effects Models.
- [2] Poirier, V., Roumet, C., Munson, A.D., 2018. The root of the matter : Linking root traits and soil organic matter stabilization processes. *Soil Biol. Biochem.* 120, 246–259. <https://doi.org/10.1016/j.soilbio.2018.02.016>
- [3] R Core Team, 2018. R: A Language and Environment for Statistical Computing.
- [4] Sokol, N.W., Sanderman, J., Bradford, M.A., 2018. Pathways of mineral - associated soil organic matter formation : Integrating the role of plant carbon source , chemistry , and point of entry. *Glob. Chang. Biol.* 1–13. <https://doi.org/10.1111/gcb.14482>
- [5] Wolfinger, R.D., 1998. Fitting nonlinear mixed models with the NLMIXED procedure. *SAS Publ.*

## **APLICACIÓN DE REGRESIÓN LOGÍSTICA DE EFECTOS MIXTOS Y VALIDACIÓN CRUZADA PARA PREDECIR LA CAÍDA NATURAL DE FRUTOS EN PERAS DEL CULTIVAR “WILLIAMS”**

GIMENEZ GUSTAVO<sup>1</sup>, RUBIO NATALIA<sup>1</sup>, BRAMARDI SERGIO<sup>1</sup>

<sup>1</sup>*Facultad de Economía y Administración, Universidad Nacional del Comahue  
gunegim@gmail.com*

### **RESUMEN**

En fruticultura se destacan labores como el raleo que permite regular la carga frutal del árbol. Esta práctica depende del número de frutos fecundados y del porcentaje de caídas que suceden naturalmente. Conocer con antelación al raleo los frutos que van a caer naturalmente y los frutos que van a persistir hasta cosecha sería clave para la toma de decisión en esta práctica. Estadísticamente, una técnica que permite clasificar en dos estados una variable de respuesta es la regresión logística y una posible herramienta para una mejor estrategia de raleo. Para ello se retomaron datos de un ensayo realizado en una parcela comercial de peras ‘Williams’ donde al inicio de la temporada se marcaron frutos en distintas plantas y se les midió el diámetro ecuatorial. Posteriormente se fueron registrando la caída del fruto y el momento en que esto ocurría. Puesto que los frutos se encuentran en estructuras del árbol y que los mismos en distintos árboles surge la necesidad de contemplar estos efectos mediante un modelo mixto. El modelo final correspondió al modelo de regresión logística considerando el diámetro del fruto y como efecto aleatorio la planta. En la pendiente del modelo estimado se encontró que la chance de caída del fruto disminuía con el diámetro del fruto. El punto de corte que define si un fruto se clasifica como caído o retenido fue de 0,927 que correspondió a un diámetro de 12,3 mm del fruto. Luego se evaluó la capacidad predictiva del modelo mediante validación cruzada y se obtuvo una precisión promedio de 0,95 en la predicción de caída de frutos. Por lo anteriormente dicho el modelo podría usarse como herramienta en la toma de decisión para el raleo de frutos en peras Williams.

**Palabras clave:** *raleo, clasificación, capacidad predictiva, curvas ROC.*

### **INTRODUCCIÓN**

En fruticultura, en la producción de peras y manzanas se destaca como tarea cultural el raleo, que consiste en la regulación de la carga frutal y permite obtener a cosecha frutos de mayores tamaños. Si bien los árboles frutales poseen mecanismos naturales de regulación como la caída natural o “June Drop” esto depende entre otros factores de la capacidad fisiológica del árbol. A pesar del “June Drop” el árbol retiene mayor número de frutos de los que son necesarios para lograr a cosecha frutos de tamaños comerciales. El raleo consiste en extraer los frutos que se encuentran en exceso de las plantas y de peores condiciones para propiciar el crecimiento y el desarrollo de los restantes. Entre las técnicas de raleo se destaca el raleo químico que consiste en aplicar un producto que genera en la planta una mayor caída de frutos que la natural. La aplicación de estos productos es una decisión delicada puesto que si se realiza cuando la carga frutal es baja, se puede generar una caída de frutos que repercuta en la cantidad total con la consiguiente pérdida de producción y un ingreso menor para el productor. Por tal motivo, sería de gran interés conocer con anticipación qué frutos permanecerán en el

árbol y que frutos caerán. Algunos trabajos [1] han encontrado una estrecha relación entre el crecimiento de los frutos y su caída durante la fase de crecimiento. Cuando la tasa de crecimiento era baja las posibilidades de caer se incrementaban. A partir de estos antecedentes se propone evaluar el diámetro del fruto lo más temprano en la temporada de crecimiento para determinar si el fruto cae o no cae de acuerdo al diámetro del mismo y estimar la carga antes de aplicar raleo alguno.

Estadísticamente, los modelos de regresión logística permiten ajustar una variable que es dicotómica, en este caso considerando la caída del fruto ( $Y=1$ ) y el desarrollo completo del fruto ( $Y=0$ ) y como variables predictoras el diámetro del fruto al momento temprano de medición. Dado que los frutos se encuentran en estructuras fructíferas como dardos y éstas en plantas se justifica la aplicación de un modelo de regresión logística con efectos mixtos[2] que contemple el efecto de la planta. Por otro lado, se esperaría que los frutos de mayor tamaño al inicio de la temporada tengan mayor probabilidad de ser retenidos que los frutos pequeños. No obstante, las regresiones logísticas predicen en término de probabilidades mientras que se requiere clasificar a los frutos como caído y no caído. Para ello se define un punto de corte ( $c$ ) a partir del cual se considera que el fruto cae si el valor de probabilidad es mayor al punto de corte o no cae si es menor al punto de corte.

La capacidad predictiva en la regresión logística se evalúa mediante las matrices de confusión que permite comparar los verdaderos positivos (TP), verdaderos negativos (TN), falsos positivos (FP) y falsos negativos (FN) en la clasificación. Y también mediante las curvas ROC (*Receiver Operating Characteristic*) que representan la relación entre la tasa de verdaderos positivos (TPR) y falsos positivos (FPR), la primera también denominada sensibilidad, evaluado en cada punto de corte. Cuanto más se aproxime la curva al eje de verdaderos positivos y mayor área bajo la curva (AUC) abarque mejor será la capacidad predictiva del modelo. El valor  $c$  se puede calcular en el punto óptimo entre la sensibilidad y la especificidad ( $TN/(TN+FP)$ ), es decir, el punto donde se maximizan ambos índices de predicción. El objetivo de este trabajo es encontrar un modelo de gran capacidad predictiva que permita conocer con antelación los frutos a caer y sea utilizado como herramienta de decisión.

## **MATERIALES Y MÉTODOS**

Se retoman datos de un ensayo realizado en el año 2008 sobre una parcela comercial de peras 'Williams' un cultivar temprano y de alto valor comercial. En dicha parcela se seleccionaron 12 plantas frutales, en cada una se demarcaron estructuras fructíferas (o dardos) sobre las cuales se identificaron todos los frutos de la misma como así también la cantidad de frutos que correspondían a la misma estructura, con un total de 774 frutos. A los 18 días después de la plena floración (ddplf) se midió en milímetros, utilizando un calibre digital, todos los frutos identificados. Posteriormente, cada 3 días y hasta la recolección comercial de los frutos, se fueron registrando los frutos caídos en cada planta y estructura fructífera. Para la caída de frutos se ajustó un modelo lineal generalizado mixto utilizando la librería lme4 del software R[3]. Para evaluar la precisión de las predicciones se realizó una matriz de confusión a partir de la librería caret. Para encontrar el punto de corte se graficó la sensibilidad y especificidad de las predicciones y el punto en que ambas se cruzaron se consideró el punto óptimo, en este caso se utilizó la librería ROCR. La evaluación del modelo se realizó aplicando una validación cruzada que consistió en dividir los datos en 10 partes y utilizar 9 partes de entrenamiento y 1 de testeo (técnica denominada 10-fold-cross validation). Para realizar la validación cruzada se programó una función de ex novo en R.

## RESULTADOS Y DISCUSIÓN

La predicción con antelación de la caída de frutos permite conocer la cantidad de frutos que se van a caer y con ello el porcentaje de frutos que quedan retenidos la planta y consecuentemente la estrategia de raleo. En este trabajo se ajustó un modelo generalizado mixto con respuesta dicotómica considerando al fruto caído como  $Y=1$  y al fruto retenido como  $Y=0$ . Al ajustar el modelo el efecto aleatorio fue significativo para la planta pero no para el dardo. Respecto de los efectos fijos se ajustó un primer modelo considerando el número de frutos por dardo y el diámetro del fruto aunque el número de frutos por dardo no fue significativo. Las estimaciones y los test realizados sobre el modelo final son las que se presentan en la Tabla 1. Como se observa en la misma las estimaciones son significativas y el efecto de la planta en el modelo es asimismo significativo. Es decir, existe una importante heterogeneidad entre las plantas en lo que respecta a la caída de los frutos que podría explicarse por: la nutrición de la planta, la floración inicial y la fecundación [1].

Tabla 1: Estimaciones de los coeficientes de los efectos fijos y aleatorios(Wald se refiere al test de Wald, RV test de razón verosimilitud).

Coeficiente(efecto)	Estimación	P-valor	Devianza	Grados de Libertad
Intercepto(fijo)	11.3788(Wald)	2.14e-11	258,8	771
Diámetro(fijo)	-0.7164 (Wald)	5.97e-08		
Planta(aleatorio)	1.915(RV)	1.809e-08		

En la Tabla 1 se destaca que la pendiente del modelo es negativa, indicando que a medida que aumenta el diámetro disminuye la chance de que el fruto caiga. El cociente de chances de la caída del los frutos obtenido a partir del exponencial del coeficiente del diámetro es de  $e^{-0,7164} = 0,44$  donde por cada milímetro que aumenta el fruto la chance que caiga se reduce un 56%. La comprobación del modelo se realizó mediante una validación cruzada seleccionando el criterio 10-fold. Dado que el modelo de regresión logística predice en término de probabilidades se calculó un valor de corte para discriminar entre frutos caídos y frutos retenidos. La capacidad predictiva del modelo se evaluó a partir de la matriz de confusión y luego se representó la curva ROC como muestra la figura 1. La línea azul punteada representa una distribución uniforme donde la clasificación sería tan débil como si se realizara al azar. La curva ROC de la figura se logra obteniendo la tasa de verdaderos positivos y de falsos positivos promedio de las 10 particiones en cada punto de corte. En este caso la TPR se incrementa hasta un punto en torno al 0,8 en que el aumento va en detrimento de la FPR; el AUC definida por la curva ROC alcanza en promedio 0,84 y se obtuvo una precisión en la clasificación de 0,95 indicando que la capacidad predictiva del modelo es muy alta. El valor de corte encontrado a partir de la curva ROC fue de 0,927, que se corresponde con un diámetro de los frutos de 12,3 mm indicando que los frutos menores a este valor seguramente caen del árbol. El modelo de acuerdo al gráfico y los estadísticos predictivos indican un modelo con muy buena capacidad predictiva del fenómeno y como herramienta factible para la toma de decisiones en el raleo.

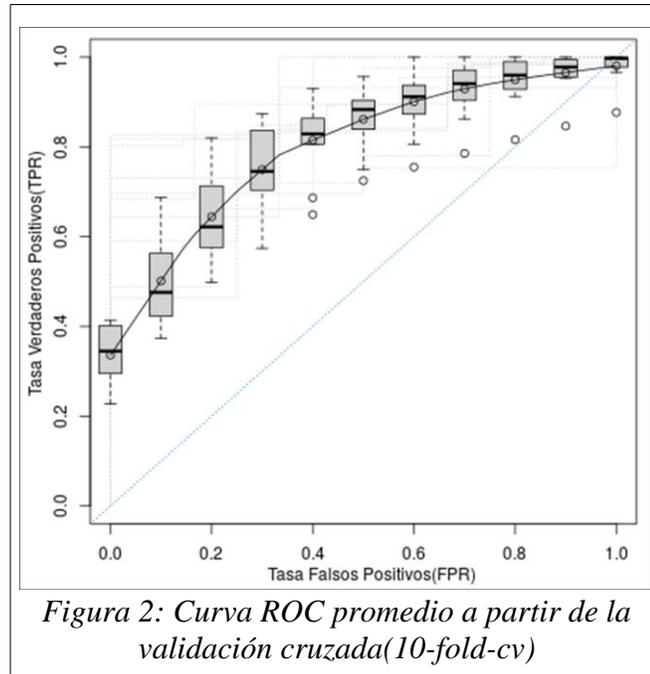


Figura 2: Curva ROC promedio a partir de la validación cruzada(10-fold-cv)

## REFERENCIAS

[1] Greene, D.;Krupa, J.;Vezinna,M.; Lakso, A.;Robinson, T.2005a. A method to predict chemical thinner response on apples. FruitNotes 70. 12-17

[2] Faraway, J. (2016). *Extending The Linear Model With R*.Second Edition. EE.UU.

[3] R Core Team (2020). A language and environment for statistical computing". Austria. R Foundation for Statistical Computing. Version 4.0.2 (2020-06-22). <https://www.R-project.org/>.

## **MODELANDO REGISTROS LONGITUDINALES DE CRECIMIENTO FÍSICO EN LA POBLACIÓN INFANTIL DE ARGENTINA: COMPARACIÓN DE MODELOS**

GONZÁLEZ GARELLO TOMÁS<sup>1</sup>, TURJANSKI PABLO<sup>1</sup>, FERNANDEZ MARÍA SOLEDAD<sup>1</sup>,  
NÚÑEZ PABLO<sup>1</sup>, PEREZ ADRIANA ALICIA<sup>1</sup>, CUETO GERARDO<sup>1</sup>

<sup>1</sup>*Grupo de Bioestadística Aplicada, (DEGE, IC, ICC) Facultad de Ciencias Exactas y Naturales (UBA-CONICET)*  
*tgarello@fmed.uba.ar*

### **RESUMEN**

El crecimiento y desarrollo infantil representa un campo de suma relevancia en salud pública. Diversas variables y/o indicadores antropométricos del crecimiento físico durante la niñez han sido fuertemente asociados a entidades crónicas en períodos posteriores a la infancia y adolescencia. El objetivo de este trabajo fue generar curvas de crecimiento en talla con cinco modelos distintos de crecimiento sujeto-específico, en función de explorar cuál tiene mejor ajuste y poder predictivo. Se trabajó con datos de niños y niñas de la población infantil vulnerable de Argentina de hasta 5 años, con al menos 10 registros de talla, provenientes de controles pediátricos asociados al programa SUMAR. Todos los modelos ajustaron correctamente, siendo el modelo de Jenss-Bayley adaptado el que mejor ajuste presentó y exhibiendo similar poder de predicción todos los modelos

**Palabras clave:** *modelos estructurados mixtos, crecimiento infantil, modelos longitudinales.*

### **INTRODUCCIÓN**

El crecimiento infantil puede considerarse como el proceso continuo a través del cual los niños y las niñas cambian en tamaño y forma en función del tiempo<sup>1</sup>. A la vez, este proceso puede ser descripto a partir del estudio de diversas medidas antropométricas a lo largo de la edad, así como la talla, el peso o la circunferencia cefálica, entre otras<sup>2</sup>. La dinámica de crecimiento infantil representa un rasgo fundamental en salud pública, dado que el comportamiento de medidas como la talla o el peso en función de la edad se encuentran íntimamente asociadas a entidades de gran relevancia pediátrica en malnutrición, así como desnutrición crónica, desnutrición aguda, sobrepeso u obesidad<sup>2</sup>. La relación entre el impacto de factores que actúan durante el desarrollo en la niñez y la salud en estadios posteriores representa un campo de estudio en crecimiento, el cual busca caracterizar y prevenir determinantes de riesgo durante el período de crecimiento temprano<sup>7,8</sup>. En este contexto, modelos matemáticos que busquen explicar o predecir trayectorias de variables antropométricas constituyen una herramienta poderosa para la evaluación clínica e investigación en salud y crecimiento infantil<sup>1</sup>. Es decir, modelos que ajusten variables que reflejen el crecimiento físico (así como talla o peso). Si bien muchos de estos modelos se usan activamente en diversas disciplinas, la literatura respecto a su aplicación en crecimiento infantil es heterogénea en términos geográficos, de tamaño de muestra, de diseño y de rangos etarios<sup>9-11</sup>, y estudios de esta índole no han sido descriptos aún para la población infantil de Argentina. En el siguiente trabajo, proponemos comparar el ajuste y poder de predicción de 5 modelos de crecimiento estructurados: “Berkey-Reeds”, “Count”,

“Gompertz Adaptado”, “Jenss-Bayley” y “Jenss-Bayley Adaptado” (BR, C, GA, JB y JBA, respectivamente) sobre registros de talla con medidas repetidas provenientes de una muestra 4857 niños y niñas de la población infantil de Argentina, desde el período postnatal hasta los 5 años.

## **MATERIALES Y MÉTODOS**

**Origen de los registros y diseño del estudio.** Los datos se originaron en consultas pediátricas de niños y niñas del programa nacional Plan Sumar (2014 - actualidad)<sup>12</sup>. Dicho programa implementado desde el Ministerio de Salud de la Nación promueve los controles de salud y educación como corresponsabilidades para mejorar las condiciones de desarrollo. Los controles de salud fueron sistemáticamente registrados y centralizados, proporcionando un gran volumen de información nacional de alta calidad y valor científico. De esta base de datos, fueron seleccionados para el presente estudio los registros correspondientes a niños y niñas de la provincia de Jujuy desde la primera semana de vida hasta los 5 años, con un mínimo de 10 mediciones de talla, identificados con un ID único y anonimizado. Luego de aplicar dichas restricciones, las bases generadas representaron en total de 2452 niñas y 2405 niños con 26584 y 26068 registros de talla respectivamente.

**Modelos de crecimiento.** Antes de modelar los datos de talla, los registros correspondientes a cada sexo fueron divididos aleatoriamente según el identificador único de cada niño o niña en una base de entrenamiento y una base de prueba en una proporción 50:50, con el fin de validar el poder predictivo de cada modelo. Cada modelo se ejecutó para la misma base de entrenamiento de cada sexo, en función de la edad como efecto fijo y contemplando a cada individuo como efecto aleatorio. El estudio comparativo se realizó en los siguientes modelos estructurales: modelo de Berkey-Reeds de orden 1, modelo de Count<sup>1,10</sup>, modelo de Jenss-Bayley<sup>1,10,11</sup>, modelo de Jenss-Bayley adaptado<sup>10</sup> y el modelo de Gompertz adaptado<sup>11</sup>. Todos los modelos fueron ajustados utilizando el paquete “nlme” del software R.

**Comparación de modelos.** El buen ajuste de los modelos evaluados fue comparado utilizando el criterio de información de Akaike (AIC), el criterio de información bayesiano (BIC) y el desvío estándar residual (RSD). Para los 3 valores obtenidos, valores más bajos indican mejor ajuste del modelo a los datos. Por otro lado, el poder predictivo de los modelos fue evaluado por el método de validación por retención. La ecuación ajustada para los efectos fijos en las bases de entrenamiento de cada sexo se utilizó para predecir el valor de cada observación de las bases de prueba del sexo respectivo. Para cada modelo y para cada sexo, se obtuvo la raíz del error cuadrático medio (RMSE) de la diferencia entre las predicciones generadas y los valores observados en la base de prueba. Valores más bajos de RMSE representaron un mejor poder predictivo por parte del modelo.

## **RESULTADOS Y DISCUSIÓN**

Los cinco modelos ajustaron los datos correctamente. Todos los parámetros pudieron ser modelados en función de las variables tanto de efectos fijos como de efectos aleatorios, a excepción de los modelos JB y JBA, en donde uno de los cuatro parámetros que tiene cada modelo no pudo ser modelado en función de los efectos aleatorios. El modelo de JBA fue el que mejor ajustó al crecimiento en talla masculino y femenino, dado que presentó los valores más bajos para los criterios de buen ajuste AIC y BIC, mientras que no mostró diferencias considerables respecto a los otros modelos en el RSD (figura 1). Por otro lado, si bien el modelo GA fue el que menor valor de RMSE arrojó, la diferencia respecto al RMSE del modelo de RMSE más bajo es de 0.06 cm para la curva de crecimiento femenina y de 0.08 cm para la

curva de crecimiento masculina, por lo que no consideramos que el poder predictivo de uno de los modelos sea mayor que otro.

Los parámetros ajustados por cada modelo pueden interpretarse en términos biológicos, representando el proceso de crecimiento que tienen las poblaciones analizadas, por lo que constituyen una buena herramienta para describir y caracterizar potenciales determinantes de riesgo en malnutrición infantil en poblaciones vulnerables de Argentina.

Modelo	DF	Crecimiento Masculino				Crecimiento Femenino			
		AIC	BIC	RSD	RMSE	AIC	BIC	RSD	RMSE
Berkey-Reeds	16	50202.03	50321.60	1.49	3.51	51388.84	51508.72	1.51	3.63
Count	11	50271.66	50353.87	1.50	3.48	51609.89	51692.31	1.54	3.61
Gompertz Adaptado	16	52466.16	52585.74	3.88	3.45	53538.93	53658.81	3.94	3.59
Jenss-Bayley	12	50948.45	51038.13	1.59	3.53	51742.48	51832.39	1.53	3.65
Jenss-Bayley Adaptado	13	49966.07	50063.23	1.49	3.49	51085.69	51183.10	1.49	3.61

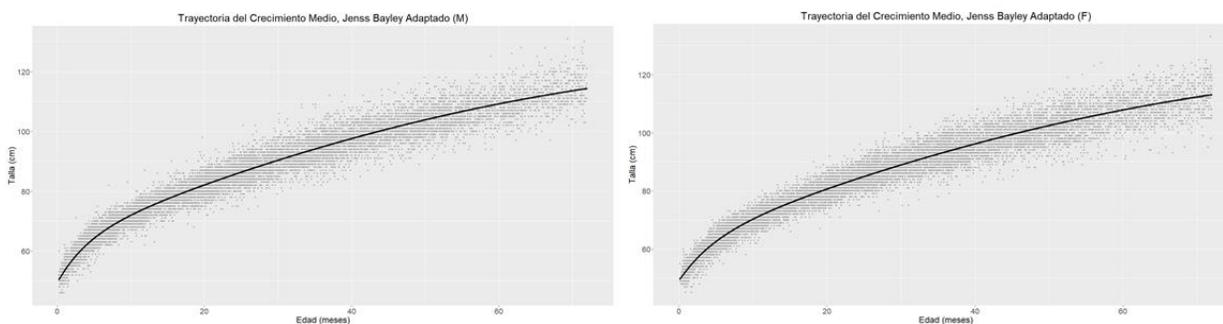


Figura 1. Panel superior: Modelos utilizados e indicadores de ajuste y predicción. (DF: número de parámetros estimados por el modelo). Panel inferior: Valores observados (puntos grises) y trayectoria del crecimiento medio en talla predicho por el modelo JBA (curva negra), para cada sexo

## REFERENCIAS

- [1] Roland Hauspie, *et al.* (2004) *Methods in Human Growth Research*. 1° Edición, Inglaterra,
- [2] Sociedad Argentina de Pediatría (2013). “Capítulo I”. Publicado en *Guía para la evaluación del crecimiento Físico*. 3° Edición, Argentina, pág 19-25
- [3] Regnault, N., & Gillman, M. W. (2014). “Importance of characterizing growth trajectories”. *Annals of Nutrition and Metabolism*, 65(2-3), 110-113
- [4] Prendergast, A. J., & Humphrey, J. H. (2014). “The stunting syndrome in developing countries. *Paediatrics and international child health*”, 34(4), 250-265
- [5] Black, R. E., Victora, C. G., *et al* (2013). Maternal and child undernutrition and overweight in low-income and middle-income countries. *The lancet*, 382(9890), 427-451
- [6] World Health Organization. (1995). “Physical status: The use of and interpretation of anthropometry, Report of a WHO Expert Committee”. World Health Organization
- [7] Heindel JJ, Balbus J, *et al.* “Developmental origins of health and disease: integrating environmental influences”. *Endocrinology* 2015; 156: 3416–21

- [8] Gillman MW, Barker D, *et al* (2007). “Meeting report on the 3rd International Congress on Developmental Origins of Health and Disease (DOHaD)”. *Pediatr Res* 61(5 pt 1):625–629
- [9] Botton J, Scherdel P, Regnault N, *et al* (2014). “Postnatal weight and height growth modeling and prediction of body mass index as a function of time for the study of growth determinants”. *Ann Nutr Metab* 65:156–66
- [10] Chirwa ED, Griffiths PL, Maleta K, Norris SA, Cameron N (2014). “Multi-level modelling of longitudinal child growth data from the Birth-to-Twenty Cohort: a comparison of growth models” (2014). *Ann Hum Biol*, 41(2): 168–179
- [11] Ahmadi *et al* (2020) “Comparison of growth models to describe growth from birth to 6 years in a Beninese cohort of children with repeated measurements”. *BMJopen*, 10(9), e035785
- [12] Nuñez, P. A., Fernández-Slezak, *et al.* (2016). “Impact of universal health coverage on child growth and nutrition in Argentina. *American journal of public health*”, 106(4), 720-726.

## DECODIFICANDO INFORMACIÓN EN REDES MULTICAPAS

HUAYLLA C.<sup>1</sup>, KUPERMAN M.<sup>2</sup>, NACIF M.<sup>1</sup>, COULIN C.<sup>1</sup>, GARIBALDI L.<sup>1</sup>

<sup>1</sup>IRNAD-CONICET, Universidad Nacional de Río Negro. San Carlos de Bariloche, Río Negro, Argentina

<sup>2</sup>Centro Atómico Bariloche (CNEA) y CONICET. Instituto Balseiro, Universidad Nacional de Cuyo, R8402AGP Bariloche, Argentina  
*cahuaylla@unrn.edu.ar*

### RESUMEN

El análisis de un ecosistema puede ser abordado usando diferentes herramientas, una de ellas son las redes. Las redes permiten capturar la estructura de la topología definida por las interacciones entre las especies. El análisis de sus propiedades constituye una forma clarificadora de entender el funcionamiento del ecosistema. Muchos trabajos no se detienen a estudiar si la red presenta información y realizan directamente el análisis de la misma. Proponemos un método para determinar si una red tiene información relevante en su sistema. Este método consiste en aleatorizar la red original intercambiando enlaces y conservando el grado de cada nodo. Si la red presenta información en su sistema, una aleatorización secuencial mostraría que los valores del índice de modularidad decrecen cuando incorporamos cambios en la red. Este método fue aplicado a una red trófica de tres niveles, polinizador, planta y herbívoro. Nosotros encontramos que la modularidad decrecía al aumentar el desorden y la frecuencia de los nodos decrece con el grado. Nuestros descubrimientos muestran que el análisis basado en la medida de modularidad es una herramienta confiable para caracterizar redes. Además, se recomienda aplicar esta metodología antes de analizar una red, ya que ésta es una herramienta para validar los resultados obtenidos.

**Palabras clave:** *ecosistema, modularidad, redes aleatorias, estructura, topología, interacción inter-específica.*

### INTRODUCCIÓN

El análisis de la interacción entre las especies en los ecosistemas es un tema central en la ecología moderna. Debido a la importancia ecológica, un considerable número de sistemas han sido relevados y sus propiedades topológicas han sido el objeto de estudio en el contexto de la teoría de redes complejas (Bascompte 2003). La estructura de una red es representada por nodos (vértices) y enlaces (conexión entre nodos). Hay resultados preliminares del efecto de la estructura de las redes de interacción en sistemas de redes abióticas, tróficas y mutualistas (Bascompte 2014; Burgos 2007). Existen además, estudios donde se analizaron la estructura de redes bipartitas, sus medidas de centralidad y la importancia de las especies (Poisot 2014). Las redes bipartitas tienen dos tipos de nodos, pueden ser representadas por dos capas, donde cada capa contiene una clase de nodo. Nos interesa ir más allá del paradigma tradicional basado en redes bipartitas mutualistas y antagonistas, entendiendo el ecosistema como mucho más que una red compleja, capaz de ser representada en varios niveles.

El objetivo es determinar si una red codifica información sobre el ecosistema. En este trabajo, consideramos que las redes ecológicas pueden contener una cantidad de información guardada en su sistema. La información es codificada por diferentes índices (intermediación, centralidad,

modularidad, etc.). Para determinar si la estructura de la red contiene información se contrasta su estructura con la de sus versiones aleatorizadas.

Si a una red se le realiza una secuencia iterativa de cambios para desordenarla y luego se calcula su índice de modularidad, debería verificarse que este índice tiene una tendencia decreciente a medida que aumenta el desorden. Este índice cuantifica la intensidad de los vínculos intra-comunidades vs. Los inter-comunidades y caracteriza la estructura de comunidades. Las comunidades son grupos disjuntos de nodos que muy probablemente comparten propiedades comunes y/o juegan un rol similar en la red. Existen diversos algoritmos que calculan este índice, nosotros usaremos el algoritmo de Louvain (Blondel 2008). Los dos puntos de nuestra investigación serán 1) determinar si una red presenta una topología que la diferencia de una red aleatoria y 2) caracterizar la distribución de grado de los nodos.

## **MATERIALES Y MÉTODOS**

Para evaluar este protocolo usamos datos relevados en un estudio sobre la relación entre polinizador-planta y herbívoro-planta. Este estudio se realizó en 'El Foyel', Rio Negro, Argentina. Usamos un subconjunto de datos a partir de Coulin (Coulin 2019) y de Nacif (Nacif 2020), los cuales corresponden a los datos de primavera 2014-verano 2015. La red fue construida con más de un nivel trófico y representada con 3 capas, una con 11 especies de plantas, otra con 11 gremios de herbívoros y otra con 25 polinizadores, con un total de 47 nodos. Se usaron las especies de plantas en común para obtener los enlaces entre capas. Se usó una matriz de adyacencia, la cual fue el resultado de proyectar la red (Boccaletti 2014).

Para lograr los objetivos recurrimos a la caracterización de redes construidas con los datos relevados y al cálculo de modularidad. Se construyó un algoritmo con el software R. El algoritmo para desordenar la red preserva el grado de cada nodo y genera redes con un número de cambios intercambiando pares de enlaces de la red original, lo que garantiza que no cambia ni el grado de los nodos ni la distribución de grado. En cada paso se considera un par de enlaces de forma aleatoria y se intercambia el nodo adyacente de cada enlace, es decir, se realiza un cambio. En caso de que la conexión esté duplicada se rechaza el movimiento. Como las redes pueden ser representadas usando una matriz de adyacencia, se considera la primera matriz como la red original. Se han considerado 10 redes con el mismo número de cambios y se ha realizado el promedio de la modularidad entre las redes que tienen el mismo número de cambios. Luego se grafica el número de cambios versus modularidad. Si hay una tendencia decreciente cuando el número de cambios se incrementa, se puede decir que la red original es diferente de una aleatoria.

Como segundo paso nos interesa calcular la distribución de grado de los nodos, ya que ésta se puede usar como un instrumento para comprender si la muestra es sesgada o incompleta. Generalmente, la distribución de grado de las redes naturales se acerca a una ley de potencia. Esta distribución es una característica de las redes biológicas (Almaas & Barabási, 2006).

## **RESULTADOS Y DISCUSIÓN**

Se observó que el valor de modularidad sigue una tendencia decreciente cuando el número de cambios se incrementa (Fig. 1). Si la red no tuviese una estructura de comunidades, no se observaría ningún patrón. La distribución de los grados de nodos muestra atributos típicos de las redes ecológicas (Fig. 2), sin embargo, debido al número reducido de nodos no sería correcto decir que se corresponde con una ley de potencia, donde la probabilidad de que un nodo tenga grado  $k$  está dada por  $P(k) \sim k^{-\gamma}$ . Recordemos que el grado de un nodo se define como la cantidad de vecinos de ese nodo.

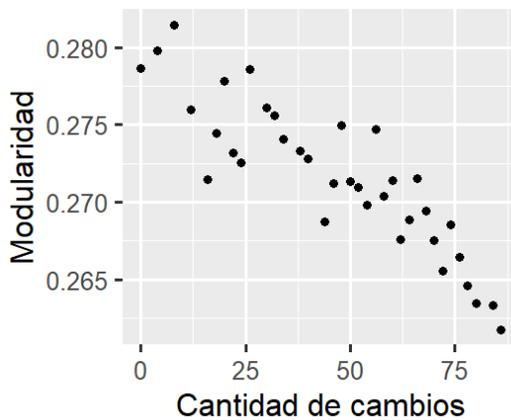


Figura 1. Modularidad vs. cantidad de cambios en la red original.

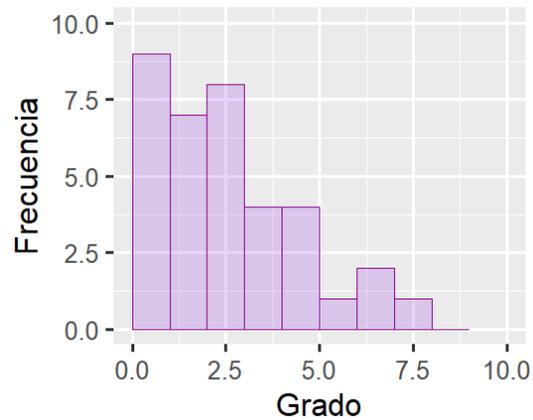


Figura 2. Histograma de la distribución de grado de los nodos

Los resultados principales son, 1) la red codifica información sobre el ecosistema, 2) la frecuencia de los nodos decrece con el grado. Se puede concluir que hubo una pérdida de información en el sistema cuando los cambios se incrementaron. Al incrementar el desorden, el sistema pierde la estructura de la red original. La diferencia entre una red natural y una aleatoria es que la red aleatoria no tiene una estructura de comunidades. Creemos que esta metodología debería ser usada antes de realizar el análisis de una red para validar los resultados obtenidos y además puede ser usado con redes de más de un nivel trófico. Como aplicación se podrán identificar especies clave en la red y características de los nodos de cada comunidad.

## REFERENCIAS

- [1] Almaas, E., y Barabási A.L. (2006) Power Laws in Biological Networks. In: Power Laws, Scale-Free Networks and Genome Biology. Molecular Biology Intelligence Unit. Springer, Boston, MA. (pp.1-11).
- [2] Bascompte, J., Jordano, P., Melian, C.J., Olesen, J.M. (2003). The nested assembly of plant-animal mutualistic networks. Proc. Natl Acad. Sci. USA, 100, 9383-9387.
- [3] Bascompte, J., and Jordano, P. (2014). Mutualistic Networks. Princeton University Press. USA.
- [4] Blondel, V. D., Guillaume, J-L., Lambiotte, R., Lefebvre, E. (2008). Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment, 2008(10).
- [5] Boccaletti, S., Bianconi, G., Criado, R., Genio, C.I, Gómez-Gardeñez, J., Romance, M., Zanin, M.(2014). The structure and dynamics of multilayer networks. Physics Reports 544 (1): 1-122.
- [6] Burgos, E. (2007). Why nestedness in mutualistic networks?. Journal of theoretical biology, 249 (2), 307-313.
- [7] Coulin, C., Aizen, M., Garibaldi L. (2019). Contrasting responses of plants and pollinators to wood-land disturbance. Wiley Online Library; Austral Ecology, 44 (6), 1040-1051.



- [8] Nacif, M., Quintero C., Garibaldi, L. One-third harvesting intensity enhances native tree performance of contrasting species while conserving herbivore diversity in a Patagonian woodland. In press.
- [9] Poisot, T., Stouffer, D., Gravel, D. (2014). Beyond species: Why ecological interaction networks vary through space and time. *Oikos*. 124 (3), 243-251.

## ANÁLISIS FACTORIAL MÚLTIPLE PARA IDENTIFICACIÓN DE ASOCIACIONES ENTRE CARACTERES FENOTÍPICOS CUALITATIVOS Y MARCADORES MOLECULARES

LAC PRUGENT V.<sup>1</sup>; COSTA TÁRTARA S.<sup>3</sup>; CURTI R.<sup>2</sup>; BRAMARDI S.<sup>1</sup>

<sup>1</sup> *Departamento de Estadística, Universidad Nacional del Comahue*

<sup>2</sup> *Cátedra de Estadística y Diseño Experimental, Sede regional Sur Metán-Universidad Nacional de Salta*

<sup>3</sup> *Departamento de Tecnología, Universidad Nacional de Luján  
lacprugentvero@gmail.com*

### RESUMEN

El mantenimiento y evaluación del material conservado en Bancos de Germoplasma son etapas indispensables para su posterior aprovechamiento. En los últimos años la caracterización agronómica y morfológica de recursos fitogenéticos se está complementando con caracterización molecular. Para contribuir a los procesos de mejora genética se espera establecer relaciones entre la caracterización agronómica-morfológica (fenotipo) de un conjunto de accesiones conservadas en un banco de germoplasma con la caracterización molecular (genotipo). En el presente trabajo, se caracterizaron 25 accesiones de quínoa nativas del Noroeste Argentino, a través de 11 variables cualitativas y 313 variables correspondientes a marcadores moleculares. Se recurrió a la técnica de Análisis Factorial Múltiple (AFM) para representar la relación entre ambas caracterizaciones. Este análisis realiza un Análisis de Correspondencias para cada grupo de variables y se define una configuración de consenso, diagonalizando la matriz yuxtapuesta, previa ponderación por el inverso del primer valor propio de cada una de las tablas que la conforman. A partir de los resultados del biplot de los dos grupos de variables cualitativas y moleculares, se observaron las posibles asociaciones, y se realizaron test chi cuadrado para testear su significancia. Se encontraron asociaciones significativas para el color de las estrías (verde, principalmente), la presencia de ramas y la posición de las ramas con la categoría sin ramas. Estos caracteres fenotipos diferenciaron a la colección estudiada con el grado de domesticación de las accesiones. Se prevé realizar, para las asociaciones significativas, modelos de mapeos asociativos. Los análisis se realizaron con el software R.

**Palabras clave:** *análisis multivariado, poblaciones nativas de quínoa, microsátélites, biplot.*

### INTRODUCCIÓN

En la búsqueda del mejoramiento genético de las especies vegetales, uno de los objetivos de los bancos de germoplasma es establecer relaciones entre la caracterización agronómica-morfológica (fenotipo) del conjunto de accesiones conservadas con la caracterización molecular (genotipo).

La quínoa es un cultivo de gran valor nutritivo que se encuentra actualmente en proceso de expansión; aprovechando sus características genéticas que le permiten generar variedades adaptables a múltiples condiciones agroecológicas. La erosión genética por pérdida de variedades de cultivos andinos ha sido preocupante en el siglo pasado [4]. Por esto es que la

conservación, mantenimiento y evaluación de material genético de la quínoa es de sumo interés.

Utilizar esta información para avanzar en el establecimiento de relaciones entre caracteres fenotípicos y los marcadores moleculares puede permitir una incipiente detección de loci asociados a estos caracteres con potencial uso en posteriores procesos de mejora [1].

Estos análisis involucran datos multivariados en los que los mismos son caracterizados por atributos cualitativos y gran cantidad de caracteres moleculares, que en el presente trabajo corresponden a regiones microsátélites denominadas en inglés Short Sequence Repeats (SSR). Los marcadores moleculares por SSR, son considerados los más robustos y polimórficos y son regiones del genoma constituidas por repeticiones en tandem de 1 a 6 pares de bases (pb) de un motivo de ADN, flanqueadas por secuencias de copia única [2].

El análisis factorial múltiple (AFM) [7], es una técnica de análisis multivariado que permite visualizar varios conjuntos de datos observados con distintas variables, a fin de encontrar una estructura común [1]. Al aplicarlo se evalúa al conjunto de accesiones, realizando dos análisis de correspondencias por separado, de estos análisis realiza una ponderación utilizando los primeros autovalores de cada uno de ellos para la construcción de la matriz yuxtapuesta [8]. Mediante un biplot que es una representación, en un mismo gráfico, de las filas (individuos) y las columnas (variables) de una matriz de datos [3], se pueden observar posibles asociaciones y testearlas luego mediante pruebas de hipótesis.

## **MATERIALES Y MÉTODOS**

En este trabajo se dispuso de una base de 25 accesiones de quínoa de la región del Noroeste de Argentina a las que se les midió 11 atributos cualitativos y se determinaron 313 variables moleculares de 22 loci SSR.

Se realizó un AFM a fin de encontrar una estructura común realizando dos análisis de correspondencias por separado.

Los atributos cualitativos fueron: Presencia de axilas (AP), Presencia de Estrías (Estrías), Color de las estrías (CEstrías), Color del tallo (Ctallo), Presencia de ramas (Rama), posición de las ramas (PorRama), Color panoja a fin anthesis (CFA), color de la panoja a cosecha (Ccos), tipo de Panoja (TP), Forma de la panoja (FP), pigmento de cotiledones (PigCot).

Se visualizaron las coordenadas de los fenotipos y los caracteres en un gráfico de biplot.

A partir de lo observado en el gráfico se testearon estas posibles asociaciones a través de Pruebas chi-cuadrado.

Para todos los análisis se utilizó el software libre R-Proyect (2009-2018 RStudio, Inc.)

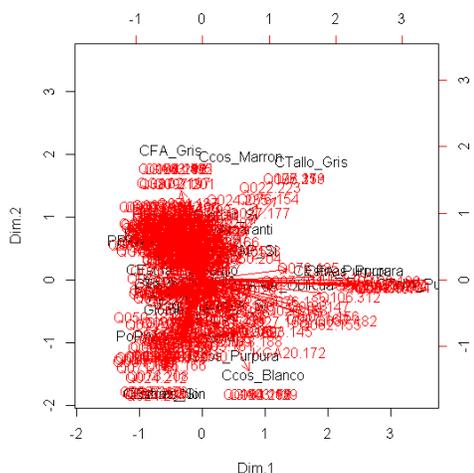
## **RESULTADOS Y DISCUSIÓN**

A partir del AFM en el cual la proporción de variancia retenida con dos dimensiones es de 21,9%, se visualizaron las coordenadas de los fenotipos y los caracteres en un BIPLLOT.

Se concluye, con un nivel de significación del 5% que el color de las estrías está asociado con los microsátélites: Q001.200, Q112.185, QGA03.153 y QCA67.208, la presencia de ramas (sí o no) y la posición de las ramas con la categoría sin ramas (sí o no) están asociados con el microsátélite Q112.185., Es importante remarcar que estos caracteres fenotípicos cualitativos (cuya herencia y expresión no está influenciada por el ambiente), son atributos que diferenciaron a la colección estudiada con respecto al grado de domesticación de las accesiones [6]. Estos resultados sustentan previos análisis realizados por Zuliani *et al.* [9], en el cual se comparan tres métodos de análisis multivariado entre los cuales está el AFM,

evidenciando una alta concordancia entre las configuraciones encontradas por los tres métodos, que se traducen en correlaciones mayores a 0,97 [9].

Las asociaciones testeadas fueron las siguientes:



Color de las estrias verde		presencia de ramas		posición de ramas sin	
Marcador molecular	p-valor	Marcador molecular	p-valor	Marcador molecular	p-valor
Q106.300	1	Q074.199	1	Q106.300	0,8063
Q027.164	0,3656	Q027.164	0,9345	Q074.199	1
Q001.200	0,0372	Q112.185	0,0372	Q027.164	0,9345
Q097.179	1	Q050.204	0,8063	Q112.185	0,0372
Q112.185	0,003538	Q106.297	0,4592	Q050.204	0,8063
Q106.297	0,1181	KGA03.151	0,8063	Q106.297	0,4592
KGA03.151	0,8063	Q024.201	0,8063	KGA03.151	0,8063
Q024.201	0,1648	Q074.199	1	Q024.201	0,8063
Q078.194	0,3656	QGA03.153	0,2113	QGA03.153	0,2113
QGA03.153	0,0372	QGA03.155	1	QGA03.155	1
QGA03.155	0,6769	QCA67.206	0,4592	QCA67.206	0,4592
QCA67.208	0,02736	QCA67.208	0,5673	QCA67.208	0,5673
Q026.182	0,6769	Q026.182	0,2113	Q026.182	0,2113

**Nota:** Los test considerados en el presente análisis fueron aquellos cuyas frecuencias esperadas eran mayores a 3, muchos de los test han sido descartados por no cumplir con esta condición.

Con base en estos resultados parciales se prevé realizar, para las asociaciones significativas, modelos de mapeos asociativos. Según los resultados obtenidos en los trabajos de Bramardi *et al.* [1] y Curti *et al.* [5] será de necesario incorporar al estudio mayor cantidad de accesiones.

## REFERENCIAS

- [1] Bramardi, S; Costa Tártara, S; Curti, R. (2017) “Métodos de Análisis Multivariado a Tres Vías Para Identificación de Asociaciones entre Caracteres Fenotípicos y Marcadores Moleculares”. XLV Coloquio Argentino de Estadística de la Sociedad Argentina de Estadística XXII Reunión Científica del Grupo Argentino de Biometría, Rosario Argentina
- [2] Sabrina M. Costa Tártara, Ramiro Curti, Carla M. Arizio, María M. Manifesto. (2013) “Quinoa del Noroeste Argentino: magnitud y estructura de la variabilidad genética del germoplasma nativo y perspectivas de uso de la colección”. Revista Ciencia y Tecnología de los Cultivos Industriales. Programa Nacional de Cultivos Industriales (Ed.). INTA. Año 3, Nº 5:32-38
- [3] Cuadras C.M (2019). *Nuevos métodos de análisis multivariante*. CMC, Barcelona España
- [4] Cruz P.; Joffre R.; Winkel T. (2015) “Racionalidades Campesinas en los Andes del Sur. Reflexiones en torno al cultivo de la quinua y otros vegetales andinos”
- [5] Curti R.; De la Vega A.; Andrade J.; Bramardi S.; Bertero H. “Multi-environmental evaluation for grain yield and its physiological determinants of quinoa genotypes across Northwest Argentina”(2014) Elsevier. <http://dx.doi.org/10.1016/j.fcr.2014.06.011>. Field Crops Research pág. 46-57
- [6] Curti R.; Andrade J.; Bramardi S.; Velásquez B., Bertero H. (2012) Ecogeographic structure of phenotypic diversity in cultivated populations of quinoa from Northwest Argentina. *Annals of Applied Biology* pág 116-125
- [7] Escofier B. and Pagès J., (1992). *Análisis Factoriales Simples y Múltiples: Objetivos,*



Métodos e Interpretación. Universidad del País Vasco. Bilbao

[8] Lavalle A; Zuliani R.; Fuentealba J.; Curti R.; Bramardi S, (2014) “Caracterización de Poblaciones Nativas de Quinoa Mediante Análisis de Datos de Conjuntos Múltiples”. *XIX Reunión Científica del GAB*, Santiago del Estero, Argentina

[9] Zuliani R.; Fuentealba J.; Lavalle A.; Bramardi S. (2013). “Análisis de los Espacios de Representación del Análisis Factorial Múltiple, Statis y Análisis de Procrustes Generalizado a partir de Una Aplicación en el Campo de la Agronomía”. *IV Encuentro Iberoamericano de Biometría y XVIII Reunión Científica del GAB*, Mar del Plata Argentina.

## EL EFECTO DE LA DISTANCIA MARÍTIMA EN EL COMERCIO DE PRODUCTOS PESQUEROS: ANÁLISIS GRAVITACIONAL

LACAZE MARÍA VICTORIA<sup>1</sup>

<sup>1</sup>Facultad de Ciencias Económicas y Sociales, Universidad Nacional de Mar del Plata  
mvlacaze@mdp.edu.ar

### RESUMEN

En las últimas décadas, se ha producido una significativa disminución de las medidas arancelarias al comercio de alimentos, junto con la expansión de estándares de calidad. En el sector pesquero, los estándares más importantes son las certificaciones ambientales, que pueden estimular el intercambio comercial o bien, operar como medidas no arancelarias. La tesis doctoral de la autora de esta ponencia, aportó la primera evidencia econométrica al respecto. A tal fin, el estudio estimó un modelo gravitacional estructural, sobre un panel de datos de comercio bilateral de productos pesqueros (Lacaze, 2018).

Esta ponencia desarrolla una contribución específica de esa investigación, la construcción de una medida de distancia entre países, alternativa a la empleada tradicionalmente en la literatura de modelos gravitacionales y su inclusión en una de las extensiones del modelo gravitacional estimado. Los resultados obtenidos dan cuenta de que el efecto negativo de la distancia, en los flujos comerciales, resulta más pequeño si se utiliza la distancia marítima, respecto del efecto que produce la distancia entre capitales. Además, los efectos positivos que, sobre el comercio, producen otros predictores, resultan mayores cuando la medida de distancia considerada es la marítima.

**Palabras clave:** *modelo gravitacional, distancia marítima, pesca, eco-etiquetado.*

### INTRODUCCIÓN

La pesca desempeña un rol clave en el sistema alimentario mundial. En los últimos sesenta años, el crecimiento anual del consumo global de pescado ha duplicado al crecimiento poblacional del planeta. La oferta alimenticia pesquera está explicada por la expansión de la acuicultura. El comercio de alimentos pesqueros también ha crecido, fundamentalmente en los últimos cuarenta años, debido a la liberalización del comercio, la globalización sectorial y mejoras en el procesamiento, la refrigeración y el transporte. Más de la mitad de las exportaciones globales de productos pesqueros proceden, desde inicios del segundo milenio, de países en desarrollo. Para muchos de estos países, dichas exportaciones constituyen la principal fuente de captación de divisas, por lo que la actividad reviste de una importancia macroeconómica estratégica.

La problemática actual de la pesca mundial, gira en torno a su sustentabilidad. Ello incluye la preocupación por su impacto ambiental, así como las consecuencias sociales y económicas, presentes y futuras. En un entorno productivo cada vez más globalizado y deslocalizado, se han generado mayores necesidades de información sobre la calidad de los insumos de la cadena de oferta. Esta información es provista a través de diversos estándares de calidad.

Las certificaciones ambientales son los estándares de calidad más importantes del sector pesquero. Establecidas mediante esquemas de eco-etiquetado, señalizan la procedencia de las capturas y la trazabilidad de toda la cadena de oferta. La información provista consolida

mercados de productos diferenciados. El eco-etiquetado puede contribuir a mejorar la gestión pesquera, a partir de una mayor demanda de productos procedentes de pesquerías sustentables. La adopción de estos esquemas puede estimular el intercambio comercial, pero también puede operar como una medida no arancelaria, generando el efecto contrario.

Una de las estrategias econométricas aplicadas para estudiar los efectos que produce en el comercio la implementación de medidas comerciales arancelarias o no arancelarias, es la llamada ecuación gravitacional. Concretamente, postula que el intercambio comercial entre países resulta proporcional a su proximidad y a los tamaños de sus respectivos mercados. Dos grupos de factores explicativos entran en dicha ecuación: los que generan atracción comercial - idioma común, adyacencia, lazos coloniales, pertenencia a acuerdos comerciales- y los que generan impedimentos -distancia geográfica, fletes, medidas de política comercial-.

La medida de distancia habitualmente empleada en la ecuación gravitacional, es la distancia bilateral entre las ciudades capitales, bajo el supuesto de que la capital coincide con el centro económico del país o se localiza cercanamente al mismo. Esta medida se encuentra tabulada para su uso en el *Centre d'Etudes Prospectives et d'Informations Internationales*.

Recientes investigaciones han objetado el valor explicativo de esta medida, para el caso particular de modelos de comercio de productos pesqueros, debido a que la producción y el comercio de estos alimentos no necesariamente se encuentran físicamente vinculados al territorio del país de procedencia (Natale *et al.*, 2015). Efectivamente, la actividad pesquera habilita la posibilidad de ubicar la flota en aguas ajenas a la zona económica exclusiva de cada país, mediante la adquisición de derechos de pesca. En consecuencia, se planteó, en la investigación doctoral referida en el Resumen, la posibilidad de construir una medida de distancia marítima o entre puertos, para su posterior inclusión en una de las extensiones del modelo gravitacional de comercio estimado.<sup>1</sup>

## MATERIALES Y MÉTODOS

Para construir la distancia marítima, se identificó, para cada país, el principal puerto de comercio internacional de productos pesqueros (Huntington *et al.*, 2015). Luego se computaron las distancias marítimas puertos de origen y destino utilizando, a tal fin, la aplicación *online SeaRates*, que proporciona las rutas de navegación comercial de uso habitual. Se empleó el Anuario de Estadísticas de Capturas y Acuicultura de la FAO, para identificar la nómina de áreas de pesca en las que se desempeñan los países y analizar la importancia relativa de cada área de pesca en el volumen total capturado por cada país exportador. Se identificaron 10 países exportadores que acusan, para el período bajo estudio, importantes proporciones de capturas en áreas no adyacentes a sus litorales marítimos.<sup>2</sup> Por lo tanto, se computaron las coordenadas geográficas promedio para las áreas de pesca en cuestión y se obtuvieron las

---

<sup>1</sup> El panel de datos de la investigación doctoral comprende las exportaciones mensuales de alimentos pesqueros verificados entre 38 países que, en términos de valor y para el período comprendido -enero de 2010 a diciembre de 2014-, concentran el 79% del comercio global de estos productos. La fuente principal de datos es la base de exportaciones mensuales UN Comtrade. La misma fue fusionada con una fuente elaborada a los fines del estudio, que permitió identificar flujos comerciales eco-etiquetados. Los resultados centrales de la investigación provienen de la estimación de un modelo gravitacional estructural, en el que las covariables tradicionales quedan absorbidas en un término de efectos fijos por pares de países, pues no poseen variabilidad temporal. La incorporación de estos efectos también absorbe la mayoría de los vínculos que pueden existir entre medidas de política comercial y el término de error, controlando su potencial endogeneidad. No obstante, para evaluar el efecto de la distancia marítima respecto de la distancia entre capitales, se estimó, a modo de extensión, un modelo con efectos direccionales, que elimina estos efectos fijos, permitiendo analizar el impacto de las covariables tradicionales.

<sup>2</sup> Australia, Chile, China, España, Fed. Rusa, Japón, Letonia, Noruega, Países Bajos y Polonia.

distancias entre dichas localizaciones y los puertos de entrada de los países de destino, corrigiendo las distancias marítimas inicialmente calculadas.

La ecuación estimada con efectos direccionales permite analizar el impacto de las covariables tradicionalmente empleadas en estos modelos y queda expresada por:

$$X_{ij,t}^k = \exp[\pi_{i,t}^k + \chi_{j,t}^k + \rho_{ij} + MCB_{ij,t}^k + MND_{j,t}^k] \times \varepsilon_{ij,t}^k \quad 3$$

## RESULTADOS Y DISCUSIÓN

La Tabla 1 muestra que el efecto negativo que la distancia impone en los flujos comerciales, resulta menos pronunciado al utilizar la distancia marítima. Además de captar fidedignamente la modalidad de transporte habitual en el comercio de productos pesqueros, esta variable contempla los recorridos efectivamente cuando las capturas proceden de áreas no adyacentes al litoral marítimo de cada exportador. Este resultado reforzaría las objeciones reportadas en la literatura, con respecto a utilizar la distancia entre capitales en la estimación de modelos gravitacionales para el comercio de productos pesqueros.

Tabla 3. Estimación agregada con efectos direccionales

Variables explicativas	Descripción	Fuente	Media/proporción	Estimación agregada con	
				Distancia entre capitales	Distancia marítima
<i>In_Distancia</i>	Distancia entre capitales / Distancia marítima entre puertos (kilómetros, en ln)	CEPII / SeaRates y elab. propia	8,32 / 8,37	-0,500*** (0,0166)	-0,480*** (0,0146)
<i>Idioma</i>	Lengua oficial común (1 = Poseen lengua común)	CEPII	1 = 13,3%	0,349*** (0,0267)	0,417*** (0,0247)
<i>Adyacencia</i>	Contigüidad geográfica (1 = Existe contigüidad)	CEPII	1 = 9,94%	0,823*** (0,0234)	1,043*** (0,0235)
<i>Lazos_Col</i>	Existencia de relaciones coloniales (1 = Existen o existieron lazos)	CEPII	1 = 8,58%	0,890*** (0,0240)	0,835*** (0,0218)
<i>Insularidad</i>	Condición de estado insular (1 = Al menos un socio es estado insular)	CEPII	1 = 34,1%	0,0184 (0,0408)	-0,213*** (0,0401)
<i>In_Flete</i>	Costo del flete marítimo por tonelada (Dólares USA por t, en ln)	SeaRates	3,50	-0,215*** (0,0143)	-0,139*** (0,0154)
<i>In_Eco-etiq</i>	Existencia de flujos comerciales eco-etiquetados (1 = Flujo eco-etiquetado)	MSC, Anuario FAO y elab. propia	1 = 8,66%	2,506*** (0,0504)	2,184*** (0,0495)
<i>ARC</i>	Pertenencia conjunta a un acuerdo regional de comercio (1 = Pertenecen a un ARC)	OMC y elab. propia	1 = 34,3%	0,00654 (0,0193)	0,0434** (0,0199)
<i>In_Arancel</i>	Arancel <i>ad valorem</i> tipo NMF establecido por el país destino (% sobre valor comercializado)	Banco Mundial	0,083	-5,296*** (0,231)	-5,837*** (0,228)
N				373,391	373,391
R <sup>2</sup>				0,976	0,978
N° de efectos incluidos				53,760	53,760

Notas: 1) Se computan, entre paréntesis, los errores estándar robustos; 2) \* p < 0,10; \*\* p < 0,05; \*\*\* p < 0,010; 3) La estimación agregada incluye efectos fijos direccionales (exportador-período-producto-variedad; importador-período-producto-variedad); 4) CEPII: Centre d'Etudes Prospectives et d'Informations Internationales, MSC: Marine Stewardship Council, FAO: Organización de las Naciones Unidas para la Alimentación y la Agricultura, OMC: Organización Mundial del Comercio, MNF: Arancel nación más favorecida. Fuente: Elaboración propia sobre salida de regresión Stata 14.1

Con respecto a las restantes covariables, los coeficientes estimados resultan significativos en términos estadísticos. Cuando se considera la distancia marítima, se observa un mayor efecto positivo del idioma en común y la adyacencia entre los socios comerciales, aunque un menor efecto positivo de la variable referida a los lazos coloniales. Esto no implica postular causalidad

<sup>3</sup> Donde  $X_{ij,t}^k$  son las exportaciones del producto k, desde el origen i al destino j, en el período t;  $\pi_{i,t}^k$  y  $\chi_{j,t}^k$  son los efectos fijos direccionales de i y j;  $\rho_{ij}$  contiene las variables referidas a distancia, idioma, adyacencia, lazos coloniales y fletes.  $MCB_{ij,t}^k$  son las medidas comerciales bilaterales e incluye la certificación con el eco-etiquetado analizado y la pertenencia a acuerdos regionales de comercio.  $MND_{j,t}^k$  hace referencia a las medidas proteccionistas no discriminatorias (aranceles "nación más favorecida"). Se estimó un sistema estructural agregado y también para las ocho posiciones arancelarias referidas a productos alimenticios pesqueros (HS12 a 4 dígitos de apertura). Por razones de espacio, en este resumen ampliado solo se presenta el modelo agregado.

alguna, toda vez que la función de costos asume una combinación aditiva de sus componentes, lo cual permite aislar sus efectos individuales.

Por su parte, el efecto del flete sobre los flujos comerciales, resulta más negativo si se utiliza la distancia entre capitales; resultado esperable ya que, el flete considerado, es marítimo. Finalmente, la presencia del programa de eco-etiquetado -objetivo central de la investigación- produce un efecto positivo en el comercio y el mismo resultaría mayor al incluir la distancia entre capitales. Los resultados obtenidos para los modelos a nivel de producto resultan robustos a los aquí discutidos, para el modelo agregado.

## **REFERENCIAS**

- [1] Huntington, T., Nimmo, F., Macfadyen, G. (2015). "Fish Landings at the World's Commercial Fishing Ports". *Journal of Ocean and Coastal Economics*, 2(1), 1-11.
- [2] Lacaze, M.V. (2018). "El rol de las medidas no arancelarias en el mercado global de alimentos pesqueros: una evaluación del desempeño de las certificaciones ambientales como catalizadores o barreras al comercio". Tesis de Doctorado, Universidad Católica Argentina. Disponible en <http://nulan.mdp.edu.ar/3152/>
- [3] Natale, F., Borrello, A., Arina, M. (2015). "Analysis of the determinants of international seafood trade using a gravity model". *Marine Policy*, 60, 98-106.

## ESTRATEGIA DE MUESTREO CON INFORMACIÓN ESPACIAL EN PRONÓSTICOS DE PRODUCCIÓN DE VID

LLERA J.<sup>1</sup>, MARTINENGO N.<sup>1</sup>, MORELLI CL.<sup>1</sup>, MARTÍNEZ E.<sup>1</sup>, MAURE E.<sup>1</sup>, RODRÍGUEZ PLAZA L.<sup>2</sup>, CARRILLO N.<sup>3</sup>, CORVALÁN, A.<sup>3</sup>; GALIOTTI, H.<sup>4</sup>; NAZRALA, J.<sup>4</sup>; SOSA, M.<sup>4</sup>

<sup>1</sup>Cát. Cálculo Estadístico y Biometría, Departamento Biomatemática y Físicoquímica,

<sup>2</sup>Cát. Topografía Agrícola, Departamento Ingeniería Agrícola

<sup>3</sup>Cát. Viticultura, Departamento Producción Agropecuaria

<sup>4</sup>Cát. Enología, Departamento de Ciencias Enológicas y Agroalimentarias

Facultad de Ciencias Agrarias- Universidad Nacional de Cuyo

jllera@fca.electrónico

### RESUMEN

Se tiene por objetivo el de proponer una estrategia de muestreo que incorpore la información del espacio para mejorar la estimación de los predictores en un pronóstico de producción parcelario en vid. Este pronóstico potencial se compara con un modelo de uso común que no utiliza información espacial de la parcela. En una primera fase, las unidades de muestreo en los llamados mapas temáticos se encuentran ubicadas en el espacio dentro de un área específica y la característica principal es que presentan autocorrelación espacial positiva. Los modelos ajustados de semivariograma aportan valiosa información sobre las distancias a tener en cuenta para el diseño muestral de una segunda fase de selección de las unidades de muestreo. Las unidades seleccionadas en la primera fase pueden a posteriori ser estratificadas con la información temprana aportada por la carga de racimos. Estos estratos presentan un nivel más homogéneo de algún condicionante que afecta la producción. Para cada estrato condicionante de valores de la covariable habrá un grupo de unidades de muestreo o plantas que se relacionan con producciones y calidades diferentes. Teniendo en cuenta estos mapas de sitio y las estratificaciones a posteriori de la carga de racimos se elabora una estrategia de muestreo probabilístico que permite pronosticar la producción. Se utilizó una parcela de Malbec de donde se tomó la información necesaria para elaborar dos estrategias de muestreo *con o sin* información previa del viñedo, donde ambas aplican un modelo de dos componentes para estimar la producción de la parcela. Se demuestra que ambas estrategias de muestreo determinaron intervalos de confianza 95% que pudieron contener a la producción real de la parcela, siendo la estrategia con información espacial la de menor error de pronóstico (9.5%) y menor longitud de intervalo. La ganancia relativa en precisión al aplicar esta estrategia fue del 37.1%.

**Palabras claves:** pronóstico de Producción, vid, muestreo.

### INTRODUCCIÓN

Se tiene por objetivo el proponer una estrategia de muestreo que incorpore la información espacial para mejorar la estimación de los predictores en un pronóstico de producción parcelario en vid. Los modelos de uso común no incorporan la información previa que puede obtenerse de una determinada parcela productiva. Es por ello, que se desea comparar estas estrategias de muestreo diferentes en modelos de dos componentes, con o sin información de la parcela. Bajo la condición de que el técnico no posee ningún tipo de conocimiento directo o

indirecto relacionado con la posible producción, es que se han estado proponiendo, en base a su aceptable error de pronóstico y menor error de estimación, el modelo de estimación del peso promedio de la producción por planta a partir de dos componentes (los promedios de cantidad total de racimos en la planta y el peso unitario de los racimos (Llera *et al.*, 2019). La modificación del rendimiento se vincula con la variabilidad de los factores productivos que se encuentran en el terreno. Esta variabilidad se produce dentro de cada cuartel. En viticultura de precisión se habla de variabilidad espacial, la cual se la puede definir como "el grado de cambio de los factores en el espacio que influyen en la calidad y rendimiento de las plantas de vid". Por eso se hace necesario incorporar la información de la variación espacial que se producen en el interior de los viñedos para aumentar la precisión de las estimaciones de los pronósticos de producción (Stanley Best, 2006). Las unidades de muestreo se encuentran ubicadas en el espacio y la característica principal es que pueden presentar autocorrelación espacial positiva. Se necesita definir modelos de pronósticos con sus correspondientes predictores y errores de estimación que tengan la particularidad de incorporar toda la información espacial. El enfoque de esta investigación es que se obtiene información previa con mapas de sitios específicos generados a partir de un muestreo sistemático con arranque aleatorio en una estructura de malla georreferenciada. La covariable que se utiliza para el mapeo es la cantidad de racimos por planta. En este contexto, la hipótesis de trabajo es la siguiente: los pronósticos de producción parcelaria con estrategias de muestreo basados en información espacial de una covariable en un momento determinado, es de mayor precisión y exactitud que los pronósticos sin información parcelaria previa

## MATERIALES Y MÉTODOS

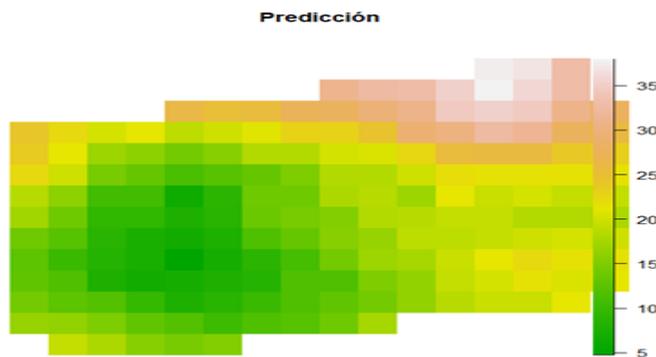
En 2019-2020, se trabajó en una parcela de la Facultad de Ciencias Agrarias. Este viñedo está compuesto por 6240 plantas de Malbec (2,14 ha), con sistema de conducción de espaldero alto compuesto de 80 hileras (a 2,40 m) donde cada una tiene 16 claros. El espaldero tiene poda en doble cordón y está orientado en el sentido Norte - Sur. En este espacio se estructuró una malla de puntos muestrales. Algunos autores proponen una cantidad mínima de 2.5 por ha, para otros sería de 15 puntos. Para poder captar la posible variabilidad espacial de las variables a medir, es que se decidió tomar una alta densidad de plantas (60 puntos.ha<sup>-1</sup>). Esto se justifica por el tamaño de muestra resultante para la variable racimos por planta, que es la de mayor variabilidad (CV%=127). Se tomó como borduras la primera y última hila, y el primer y último claro de cada hileras. La primera hileras de la estructura fue sorteada (1 a 5), al igual que la primera planta de la hileras (1 a 11). Luego en forma sistemática cada 5 hileras (12 m aproximadamente) y 8 plantas en la hileras (promedio de 15.1 m entre plantas en la hileras), se obtiene la malla de 128 puntos. Una estrategia de muestreo incluye el diseño del muestreo y la definición de los estimadores que se asocian al mismo.

**Estrategia 1:** En la primera fase del muestreo, se cuentan los racimos por plantas en la malla de puntos definida con el muestreo sistemático con arranque aleatorio y se estima su promedio. Luego en un segundo muestreo antes de cosecha, que es considerado independiente del primero, se aplica un muestreo por conglomerados con dos etapas. Las unidades de muestreo primarias (UMP) son las plantas y las unidades secundarias de muestreo (USM) son los racimos, los cuales se pesan individualmente (kg.rac<sup>-1</sup>). Con esta estrategia no se utiliza información de la posible variación espacial que afecte las mediciones. **Estrategia 2:** Con la información obtenida en la primera fase se aplica una metodología de sitio específico. Luego se realizó un muestreo estratificado proporcional según las dimensiones de cuatro sectores homogéneos. Las plantas dentro del estrato se seleccionan con muestreo aleatorio y con la

condición de que la distancia espacial entre plantas (o rango práctico) sea mayor a la indicada por el semivariograma (variación espacial).

## RESULTADOS Y DISCUSIÓN

En la Imagen 1 se puede observar cierta tendencia positiva ( $b_1=0.31$  y  $p=0.001$ ) de Oeste a Este (Entre Hileras), como de Sur a Norte (dentro de la Hileras). Hay una leve tendencia a mayor carga en las plantas ubicadas hacia el sector noreste. El cálculo del índice de Moran resultó significativo de que existe auto-correlación espacial positiva ( $IM=0.10$   $p=0.004$ ). El semivariograma teórico ajustado fue exponencial: Nugget ( $C_0=329.85$ ), Sill ( $C=177.2$ ) y Range (67.2 m). El rango práctico de  $R_p=201,6$  m es la distancia entre dos puntos en que se pueden considerar independencia de las observaciones con respecto a la varianza espacial (Balzarini, 2014). Una medida del grado de estructuración espacial, es la varianza estructural relativa ( $RSV=34.9\%$ ,  $RSV = C / (C+C_0) \times 100$ ). En este caso el RSV es *medio*. Por lo tanto, las estimaciones corrigiendo el efecto espacial deberían mejorar las eficiencias de las estimaciones (Zimback, 2001). A partir de estos parámetros estimados del modelo exponencial en el semivariograma, se realizó la interpolación mediante Kriging en bloque y se obtuvo un mapa de contorno para la variable cantidad de racimos (X19). La dimensión del bloque utilizada fue de  $20 \times 20$  m. El mapa generado muestra la variabilidad espacial (Imagen 1). La zona más oscura del color verde se corresponde con el área de menor producción unitaria de uvas en la parcela (Oeste y Sur de la parcela). Por lo contrario el sector de mayor producción estaría en el sector Noreste. Con esta información se formaron 4 estratos: 1) 0 a 20, 2) 21 a 40, 3) 41 a 60 y 4) mayor a 60 racimos. Estos se corresponden con 4 categorías de carga: Mala, Regular, Buena y Muy Buena, respectivamente.



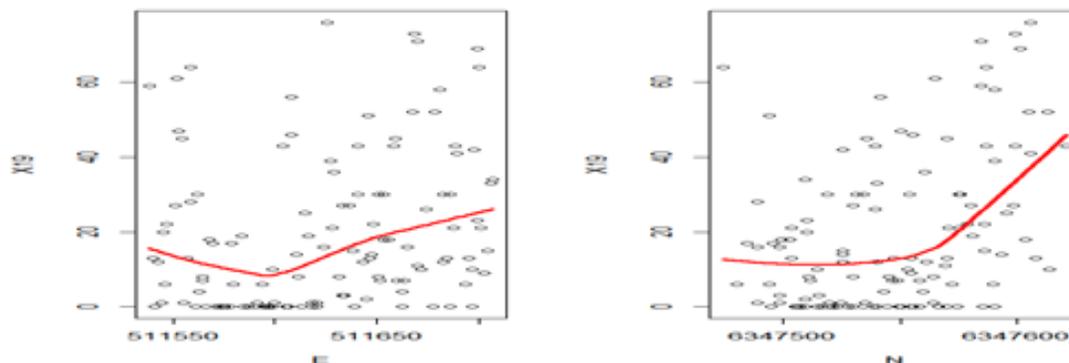


Imagen 1: Mapa de variabilidad espacial con bloques de 20x20 m, modelo exponencial para la cantidad de racimos de plantas y tendencia espacial en la parcela de vid (InfoStat)  
 En Tabla 1 se muestra que ambas estrategias de muestreo determinaron intervalos de confianza 95% que pudieron contener a la producción de la parcela (5073 kg), siendo la estrategia 2 la de menor error de pronóstico (9.5%) y menor longitud de intervalo (2276.8 kg). La ganancia relativa en precisión al aplicar esta estrategia comparada con la 1, fue del 37.1%

**Tabla 1:** Pronósticos de producción, errores de estimación y error de pronóstico para la **estrategia sin información (1)** y **con** incorporación de **información** espacial de la parcela **(2)**

Estrategia	Pronóstico (kg/parcela)	Error Estimación	Intervalo de Confianza 95%		Cosecha (kg/parcela)	Longitud (%)	% Error Pronóstico
			Límite Inferior	Límite superior			
1	5722.7	732.4	4287.2	7158.2	5073.1	2871.0	12.8
2	5557.4	580.8	4419.0	6695.8		2276.8	9.5

Con la variable cantidad de racimos por planta, valorada en forma temprana para los puntos de muestreo del lattice (primera fase), se pueden construir mapas que permiten detectar zonas homogéneas de carga de racimos. Luego se podrán formalizar los estratos para un muestreo aleatorio estratificado que permita estimar el promedio del peso unitario de los racimos (segunda fase). Esta metodología permite obtener pronósticos de producción parcelaria con estrategias de muestreo basados en información previa de mapas de covariables con referencia espacial para un momento determinado, que permiten mayor precisión y exactitud que los modelos de pronósticos sin información parcelaria previa. Con la información de los mapas de sitio también se podrían realizar cosechas con calidades enológicas diferentes.

## REFERENCIAS

- [1] Balzarini, M.; Córdoba, M.; Bruno, C.; Aguade, F.; Tablada M. (2014) *Análisis de la variabilidad espacial en lotes agrícolas. Manual de Buenas Prácticas Agrícolas- 1a ed.* – Córdoba: Eudecor, 2014. 119 p.; 25x17 cm. ISBN 978-987-1536-66-5
- [2] Llera, J.; Galiotti, H.; Martinengo, N.; Matus, S. (2017). *Muestreo probabilístico en viñedos de uvas tintas para estimar la calidad del mosto.* ISSN: 0370-4661 – Revista de la Facultad de Ciencias Agrarias-UNCuyo. Tomo 49 Nro 1- 119-126- SECTyP. UNCuyo.
- [3] Llera, J.; Martinengo, N.; Morelli, C.; Martínez, E.; Maure Russo, E.; Rodríguez Plaza, L.; Carrillo, N.; Corvalán, A.; Galiotti, H. (2019) *Pronósticos de producción en Vid utilizando predictores con información espacial.* XXIV Reunión Científica del GAB. FCA, UNCUIYO, Mendoza, Argentina.

## **MEDICINA BASADA EN LA EVIDENCIA: PRUEBAS DIAGNÓSTICAS Y SU APLICACIÓN EN ENSAYOS DE EVALUACIÓN DE DESEMPEÑO DE UN TEST RÁPIDO DE DETECCIÓN DE COVID 19**

LOPEZ VANESA<sup>1</sup>, DELFINO VANINA<sup>1</sup>, FILIPPINI O. SUSANA<sup>2</sup>

<sup>1</sup>Universidad Nac. de Tres de Febrero, Depto. de Estadística y Matemática

<sup>2</sup>Universidad Nacional de Luján, Depto. de Ciencias Básicas  
vanecele@gmail.com

### **RESUMEN**

La Medicina basada en Evidencias (MBE) involucra el uso consciente y explícito de las mejores pruebas actuales en la toma de decisiones sobre la atención de cada paciente; comprende la realización de diagnósticos efectivos y la consideración de los principios y los derechos de los pacientes. A partir de MBE, se pueden abordar distintos aspectos, cómo determinar causas de enfermedades y diagnósticos, a partir de baterías de *Pruebas Diagnósticas*, valorando su certeza y precisión. Para un buen diagnóstico es importante, evaluar en forma crítica la validez de la evidencia, impacto y aplicabilidad. La investigación clínica centrada en el paciente incluye análisis sobre exactitud de pruebas diagnósticas, marcadores pronósticos y eficacia y seguridad de las pautas terapéuticas, preventivas y de rehabilitación. Una *prueba diagnóstica* (PD) es cualquier proceso que pretenda determinar en un paciente la presencia de cierta condición, supuestamente patológica. Una buena prueba, ofrece resultados positivos en enfermos y negativos en sanos. Las condiciones que deben ser exigidas en una prueba son tres: Validez (sensibilidad y especificidad), Reproductividad y Seguridad (valor predictivo positivo o negativo). El objetivo de este trabajo, es evaluar a partir de aplicaciones de probabilidad, las condiciones de desempeño de un test rápido (Rt) para detectar Covid 19 vs PCR, en diferentes ensayos. Se realizó una validación en campo en 11 centros de referencia. El Test presenta una alta sensibilidad clínica: 94,7% y 96,3 % de Exactitud. La probabilidad de que un individuo con un resultado negativo en la prueba esté realmente sano es 0,89.

**Palabras clave:** *test rápido, sensibilidad, especificidad, exactitud diagnóstica.*

### **INTRODUCCIÓN**

Para muchos médicos, la "evidencia" es a menudo una combinación vaga de estrategias eficaces recolectadas en pacientes anteriores, el asesoramiento de mentores y colegas, y una impresión de "qué se está haciendo" sobre la base de artículos, resúmenes, simposios y anuncios seleccionados al azar. La Medicina basada en Evidencias (MBE), involucra el uso consciente, explícito y juicioso de las mejores pruebas actuales en la toma de decisiones sobre la atención de cada paciente, no sólo sobre su enfermedad o proceso. Ello comprende desde la realización de diagnósticos efectivos y eficientes, a la identificación y consideración de los principios, las preferencias y los derechos de los pacientes. A partir de MBE, se pueden abordar distintos aspectos, entre ellos, hallazgos clínicos, etiologías, o cómo determinar causas de enfermedades y diagnósticos, a partir de Pruebas o baterías de Pruebas Diagnósticas, determinando la valoración de la certeza y precisión de las mismas. Para un buen diagnóstico, evaluar en forma crítica la validez de la evidencia (aproximación a la verdad), impacto (tamaño del efecto) y aplicabilidad (utilidad en la práctica cotidiana), es importante. En investigación

clínica centrada en el paciente, por mejores evidencias debe entenderse lo que incluye análisis sobre exactitud de pruebas diagnósticas, marcadores pronósticos y eficacia y seguridad de las pautas terapéuticas, preventivas y de rehabilitación. Se llamará prueba diagnóstica (PD) a cualquier proceso, más o menos complejo, que pretenda determinar en un paciente la presencia de cierta condición, supuestamente patológica, no susceptible de ser observada directamente. Es evidente que una buena prueba diagnóstica es la que ofrece resultados positivos en enfermos y negativos en sanos (Miettinen, O. (2011) *Epidemiological Research Terms And Concepts*). Por lo tanto, las condiciones que deben ser exigidas a un test son tres: Validez: es el grado en que un test mide lo que se supone que debe medir. ¿Con que frecuencia el resultado del test es confirmado por procedimientos diagnósticos más complejos y rigurosos? La sensibilidad y la especificidad de un test son medidas de su validez. Reproductividad: es la capacidad del test para ofrecer los mismos resultados cuando se repite su aplicación en circunstancias similares. Seguridad: viene determinada por el valor predictivo de un resultado positivo o negativo. ¿Con que seguridad un test predecirá la presencia o ausencia de enfermedad? El objetivo de este trabajo, es mostrar cómo a partir de sencillas aplicaciones de probabilidad se evalúan las condiciones de desempeño del Covid 19 NEOKIT (Rt), en diferentes ensayos versus PCR.

### MATERIALES Y MÉTODOS

A partir de información de investigadores del ICT MILSTEIN – CONICET, se analizó el nuevo kit de diagnóstico de COVID-19 (Rt), que fuera validado por la Administración Nacional de Medicamentos Alimento y Tecnología Médica (ANMAT). Se trata de un kit de bajo costo y fácil de maniobrar, que permite indicar en menos de dos horas, a partir de una muestra respiratoria, si una persona está o no infectada por el Coronavirus SARS-CoV-2. Dado que es un test rápido, permite fácil aplicación y pronto diagnóstico. Una **prueba diagnóstica** (PD) es cualquier proceso que pretenda determinar en un paciente la presencia de cierta condición, supuestamente patológica. Una buena prueba, ofrece resultados positivos en enfermos y negativos en sanos. Es una prueba dicotómica que clasifica cada paciente con resultado negativo o positivo. Los conceptos relacionados a PD generan dentro del espacio, 4 casos de probabilidades condicionales, que surgen del análisis de una tabla de contingencia. Ver Tabla 1. En la tabla de contingencia se enfrenta el resultado de la prueba diagnóstica (Rt) (filas) con el estado real de los pacientes o el resultado de la prueba de referencia o «gold estándar» utilizada (PCR) (columnas). Se analizan la sensibilidad y la especificidad para evaluar la validez de la prueba, y los valores predictivos positivos y negativos para evaluar la seguridad.

Tabla 1: Tabla de contingencia para el estudio de validez y exactitud diagnóstica.

Resultado de la Prueba	ESTADO VERDADERO		Total
	Enfermo	Sano	
Positivo	Verdaderos Positivo(VP)	Falsos Positivos (FP)	VP + FP
Negativo	Falsos Negativos (FN)	Verdaderos Negativo(VN)	FN+VN
TOTAL	VP + FN	FP +VN	VP+VN+FP+FN

La sensibilidad es la probabilidad de que para un sujeto enfermo se obtenga en la prueba un resultado positivo.

La especificidad es la probabilidad de que para un sujeto sano se obtenga un resultado negativo.

El valor predictivo positivo (VPP) es la probabilidad de padecer la enfermedad si se obtiene un resultado positivo en el test. El valor predictivo negativo (VPN) es la probabilidad de que un sujeto con resultado negativo en la prueba esté realmente sano.

La exactitud es la probabilidad de que el test prediga correctamente la presencia o ausencia de la enfermedad.

$$Sensibilidad = \frac{VP}{VP+FN} \quad Especificidad = \frac{VN}{VN+FP} \quad VPP = \frac{VP}{VP+FP} \quad VPN = \frac{VN}{FN+VN}$$

## RESULTADOS Y DISCUSIÓN

Se observa, analizando el ensayo Nro. 1 de la tabla a modo de ejemplo, que la proporción de individuos infectados que fueron correctamente diagnosticados con la enfermedad es 0,94, o sea que presenta alta sensibilidad clínica. Asimismo, el porcentaje de individuos sanos que fueron correctamente diagnosticados por la prueba es 100%. La probabilidad de que el individuo esté enfermo de COVID-19 dado que el resultado del test es positivo, es igual a 1 (VPP). La probabilidad de que un individuo con un resultado negativo en la prueba esté realmente sano es 0,89 (VPN). La probabilidad de que el test prediga correctamente la presencia o ausencia de la enfermedad es 0,96.

Se realizó una amplia validación en campo con 562 muestras. Rt versus-PCR en 11 centros de referencia. El Test presenta una alta sensibilidad clínica: 94,7%. (No se reproduce en su totalidad la investigación, por falta de la totalidad de los ensayos, ya que hay informes en proceso).

### Análisis de desempeño del Test de diagnóstico molecular simplificado del virus SARS CoV-2 Rt VS Test -PCR.

Ensayo* n°	Cantidad de Muestras	Sensibilidad (%)	Especificidad (%)	Exactitud diagnóstica (%)
1	150	94,0	100,0	96,0
2	82	97,1	100,0	98,8
3	92	87,9	96,6	93,5
4	21	100,0	100,0	100,0
5	20	95,0	**	95,0
6	62	90,9	97,5	95,2
7	40	100,0	91,2	92,5
8	15	100,0	100,0	100,0
9	28	94,4	100,0	96,4
10	42	100,0	100,0	100,0
11	10	100,0	100,0	100,0
	<b>562</b>	<b>94,7</b>	<b>97,9</b>	<b>96,3</b>



(\*) Incluye ensayo para aprobación del Test por ANMAT (muestras provistas por ANLIS MALBRÁN), y pruebas de desempeño en campo. Hospitales San Juan de Dios (La Plata, PBA.), Lab. Salud Pública (FCE, UNLP), Rossi (La Plata, PBA.), Muñiz (CABA), ANLIS MALBRÁN (CABA), Federico Abete (Malvinas Arg., PBA.), Cordero (San Fernando, PBA.), Elizalde (CABA), N.Kirchner (Escobar, PBA.), Santojanni (CABA).

(\*\*) No corresponde el cálculo por realizarse ensayo solo con muestras de pacientes diagnosticados positivos.

## **REFERENCIAS**

- [1] Miettinen, O. (2011) *Epidemiological Research Terms And Concepts*. Ed. Springer, Canada.
- [2] Jama. (1992). "Evidence-Based Medicine Working Group. Evidence-based medicine. A new approach to teaching the practice of medicine". 268: 2420–5.

## **PATRONES ESPACIALES EN INDICADORES DE MALNUTRICIÓN INFANTIL DE LA REPÚBLICA ARGENTINA**

MELCONIAN M.A.<sup>1</sup>, FERNANDEZ M.S.<sup>1</sup>, TURJANSKI P.<sup>1</sup>, ALTSZYLER E.<sup>1</sup>, LIBMAN M.<sup>2,3</sup>, PEREZ A.<sup>1</sup>, NUÑEZ P.<sup>1</sup>, CUETO G.R.<sup>1</sup>

<sup>1</sup> *Grupo de Bioestadística Aplicada, (DEGE-DC-IC-ICC)-FCEN-UBA/CONICET*

<sup>2</sup> *Universidad Nacional Tres de Febrero (UNTREF)*

<sup>3</sup> *Instituto del Conurbano, UNGS*

*agustinamelco@gmail.com*

### **RESUMEN**

Combatir la malnutrición infantil por exceso representa uno de los desafíos más serios en materia de salud pública del siglo. Esta problemática se encuentra fuertemente asociada con comorbilidades tanto en la infancia como en la vida adulta. En el presente trabajo se estudiaron indicadores de malnutrición infantil (sobrepeso y obesidad) de niños y niñas entre 6 y 9 años, recabados en 5364 centros de salud de la Argentina. Se implementaron modelos estadísticos geoespaciales con métodos de estimación bayesiana para estudiar la distribución espacial de las prevalencias de sobrepeso y obesidad. Se lograron identificar parches y regiones con peores situaciones nutricionales para los indicadores estudiados. La salud pública de precisión con modelado de datos geoespaciales se está convirtiendo en una herramienta esencial para evaluar el progreso y las áreas que exigen mejoras, aportando conocimiento valioso para contribuir al mejoramiento de las condiciones de crecimiento y desarrollo infantil en nuestro país.

**Palabras clave:** *malnutrición infantil, INLA, análisis espacial.*

### **INTRODUCCIÓN**

La Organización Mundial de la Salud (OMS) define el sobrepeso y la obesidad como “una anormal o excesiva acumulación de grasa que representa un riesgo para la salud”. Varios estudios han demostrado que la obesidad en la infancia se encuentra asociada con un mayor riesgo de padecer obesidad en la edad adulta, morir prematuramente o tener algún tipo de discapacidad<sup>1</sup> y distintas comorbilidades infantiles como la diabetes tipo II o afecciones a nivel respiratorio<sup>2</sup>. La prevalencia del sobrepeso como de la obesidad entre niños, niñas y adolescentes ha aumentado en una tasa alarmante durante los últimos años<sup>1</sup>. Hoy en día, combatir la obesidad infantil representa uno de los desafíos más serios en materia de salud pública del siglo. En Argentina, una combinación de programas de salud, entre ellos el Programa Sumar [2013-actualidad]<sup>3,4</sup> dependientes del Ministerio de Salud de la Nación (MSAL), se implementaron con el objetivo de reducir las desigualdades en el acceso a la salud y mejorar las condiciones de vida, con foco en las poblaciones más vulnerables. La información obtenida por el programa ha sido sistemáticamente registrada y centralizada a lo largo del tiempo, proporcionando un gran volumen de información subnacional de alta calidad y valor científico. La ciencia de datos en salud es un área de investigación emergente que promete contribuir al diseño de políticas públicas<sup>5,6,7</sup>. La ‘salud pública de precisión’ como estrategia para describir, monitorear y crear sistemas de alertas para el seguimiento dinámico de indicadores a

escalas locales (subnacionales) donde se implementan y diseñan políticas públicas, es fundamental para evaluar progresos y áreas que requieren mejoras<sup>8</sup>

En el presente trabajo, se propone analizar la variación espacial de la prevalencia de obesidad y de sobrepeso de niños y niñas de poblaciones vulnerables de toda la Argentina durante el año 2019.

## **MATERIALES Y MÉTODOS**

Se trabajó a partir de registros antropométricos del primer y segundo cuatrimestre del 2019, de niños y niñas de 6 a 9 años beneficiarios del programa SUMAR correspondientes a 5364 centros de salud (efectores) de Argentina. Para cada efector se calculó la cantidad de beneficiarios con sobrepeso u obesidad siguiendo el criterio propuesto por la OMS: individuos de entre 5 y 19 años presentan sobrepeso u obesidad si su índice de masa corporal -IMC = peso(kg)/(altura(m))<sup>2</sup>- supera en más de una o dos desviaciones estándar respectivamente a la mediana para su edad establecida en los patrones de crecimiento infantil de la OMS. Además se calculó para cada efector el total de beneficiarios que realizaron controles de salud, la edad promedio y la proporción de varones. Se aplicaron modelos geoespaciales con métodos de estimación bayesianos (Bayesian space-model-based geostatistics using integrated nested Laplace approximations), disponibles en la librería R-INLA<sup>9</sup> con el software R<sup>10</sup>. Para el modelado de la estructura espacial se utilizaron Ecuaciones Estocásticas Parcialmente Diferenciales o SPDE (Stochastic Partial Differential Equation), una malla (“mesh”) conteniendo las coordenadas de cada uno de los centros de salud (Figura 1.a) y una matriz de correlación de Mátern. Se implementaron dos modelos, uno para prevalencia de sobrepeso y otro para prevalencia de obesidad, utilizando una distribución Binomial. Se incluyeron como variables de control la proporción de varones y la edad promedio (días) para cada centro de salud.

## **RESULTADOS Y DISCUSIÓN**

La cantidad de niños/as por efector presentó un mínimo de 2 y un máximo de 1993 individuos, con media de 34 y una mediana de 63 niños/as por efector. La mediana de la prevalencia de sobrepeso (PS) y obesidad (PO) para ambos sexos fueron de 19,5% (media: 19,8%) y 17,5% (media: 18,6%) respectivamente.

Los modelos ajustados para la PO y la PS presentaron una estructura espacial significativa ( $\Delta$ DIC al incorporar el factor espacial: 6006,9 y 3683,5 respectivamente). Las prevalencias estimadas para el 90% de los efectores variaron entre [13,09-24,97]% para el sobrepeso y [9,36-26,23]% para la obesidad. Tanto la PS como la PO aumentaron significativamente con la edad promedio, y la PO aumentó además con la proporción de varones. Estos resultados se encuentran en línea con lo observado por el MSAL y UNICEF Argentina (2018)<sup>11</sup>, donde se observó un progresivo aumento de la PS Y PO con la edad y una mayor PO en los varones.

Se observó una estructura de “parches” y regiones (grupos de provincias) con valores altos y bajos de PS y PO, con casos extremos en provincias como Santiago del Estero y Buenos Aires para la PS y el sur de Buenos Aires, Río Negro y Córdoba para la PO (Figura 2 a. y b.). Para ambos indicadores, las provincias patagónicas mostraron mayores valores, lo cual nuevamente coincide con una tendencia ya observada en la publicación citada anteriormente<sup>11</sup>

Como conclusión, los modelos ajustados permitieron describir la estructura espacial presente en los indicadores de sobrepeso y obesidad e identificar de manera rápida parches y/o regiones con peores situaciones nutricionales. Teniendo en cuenta que la problemática de la malnutrición por exceso en Argentina en niños, niñas y adolescentes ha tenido un crecimiento sostenido a lo largo de los años<sup>11</sup> es de suma importancia contar con herramientas que permitan enfocar

posibles intervenciones sanitarias acordes a la situación que se presente, en grupos determinados o espacios geográficos particulares.

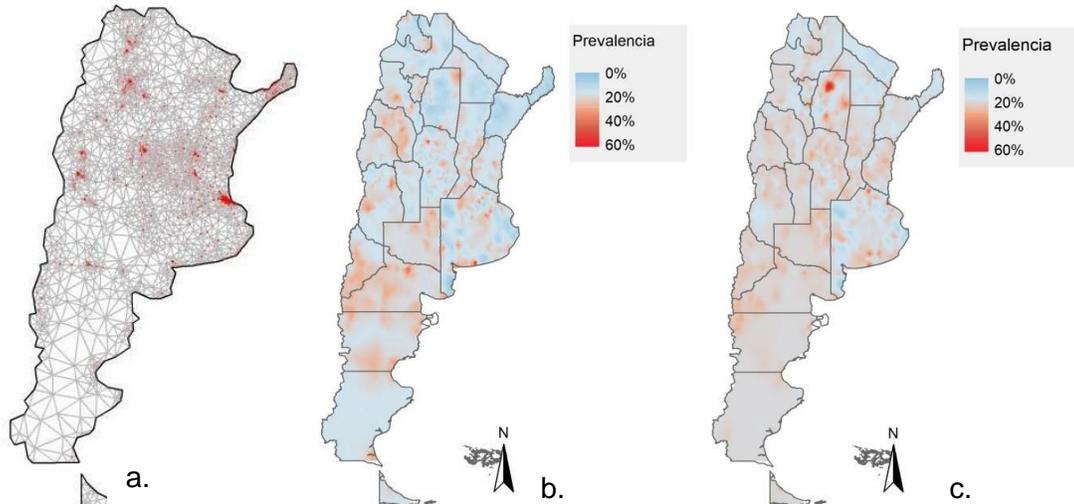


Figura 1. (a) Malla utilizada para modelar las prevalencias de obesidad y sobrepeso. En rojo se indica la localización de los centros de salud. (b), (c) Mapas elaborados con las estimaciones de los modelos para las prevalencias de obesidad (b) y sobrepeso (c)

## REFERENCIAS

- [1] World Health Organization, 2014. Global Strategy on Diet, Physical Activity and Health. . 2014.
- [2] Reilly, J J, Methven, E, Mcdowell, Z C, Hacking, B, Alexander, D, Stewart, L And Kelnar, C J H, 2003. Health consequences of obesity. 2003. P. 748–752.
- [3] Atun, Rifat, De Andrade, Luiz Odorico Monteiro, Almeida, Gisele, Cotlear, Daniel, Dmytraczenko, Tania, Frenz, Patricia, Garcia, Patricia, Gómez-Dantés, Octavio, Knaul, Felicia M, Muntaner, Carles And Others, 2015. Health-system reform and universal health coverage in Latin America. *The Lancet*. 2015. Vol. 385, no. 9974, p. 1230–1247.
- [4] Frenk, Julio, 2015. Leading the way towards universal health coverage: a call to action. *The Lancet*. 2015. Vol. 385, no. 9975, p. 1352–1358.
- [5] Murdoch, Travis B and DETSKY, Allan S, 2013. The inevitable application of big data to health care. *Jama*. 2013. Vol. 309, no. 13, p. 1351–1352.
- [6] Bellazzi, Riccardo, 2014. Big data and biomedical informatics: a challenging opportunity. *Yearbook of medical informatics*. 2014. Vol. 23, no. 01, p. 8–13.
- [7] Khoury, Muin J and Ioannidis, John P A, 2014. Big data meets public health. *Science*. 2014. Vol. 346, no. 6213, p. 1054–1055.
- [8] Dowell, Scott F, Blazes, David And Desmond-Hellmann, Susan, 2016. Four steps to precision public health. *Nature News*. 2016. Vol. 540, no. 7632, p. 189.
- [9] Blangiardo, Marta And Cameletti, Michela, 2015. *Spatial and spatio-temporal Bayesian models with R-INLA*. John Wiley & Sons.
- [10] Team, R D C, 2020. *R: a language and environment for statistical computing*. 2011. Vienna: R Foundation for Statistical Computing.
- [11] Ministerio De Salud De La Nación Argentina and UNICEF (2018). Sobrepeso y obesidad.

## ANÁLISIS EXPLORATORIO DE LAS CONDUCTAS ALIMENTARIAS: FACTORES QUE INCIDEN EN LAS ELECCIONES DE LOS ESTUDIANTES UNIVERSITARIOS<sup>4</sup>

MENÉNDEZ LUCIANO, BERGES MIRIAM, LUPÍN BEATRIZ

*Grupo de Investigación Economía Agraria, FCEyS-UNMdP*

*luciano\_menendez001@hotmail.com; mberges@mdp.edu.ar; beatrizlupin@gmail.com*

### RESUMEN

Las conductas alimentarias son de carácter complejo y origen multifactorial. Se van construyendo y modificando en la medida que las personas interactúan y se integran a la sociedad por medio de diversas instituciones: la familia, el trabajo, los vínculos o amistades, la educación, etc. En tal sentido, al ingresar a la educación superior, la estructura de saberes –y de elecciones– se enfrenta a nuevas realidades y personas, lo que puede derivar en un cambio sustancial de las conductas. Dado que la adopción de un estilo de vida saludable, preferentemente durante la juventud, es relevante para prevenir determinadas enfermedades, en los últimos años, se ha incrementado la tendencia hacia una alimentación sana y nutritiva. Este trabajo explora la incidencia de factores vinculados a las preferencias y a las percepciones respecto al grado de independencia con que estudiantes universitarios toman sus decisiones alimentarias. Se presentan los resultados obtenidos al aplicar un Análisis de Correspondencias Múltiples a datos provenientes de una encuesta en la que participaron 229 estudiantes de la Facultad de Ciencias Económicas y Sociales-Universidad Nacional de Mar del Plata (FCEyS-UNMdP).

**Palabras clave:** *análisis estadístico multivariado, consumo de alimentos, alimentación saludable, atributos de calidad, comportamiento del consumidor.*

### INTRODUCCIÓN

Al ingresar a la educación superior, los individuos conocen y actúan en nuevos lugares, estableciendo nuevos vínculos personales, profesionales y/o académicos. Así, ponen a prueba sus saberes y, en consecuencia, refuerzan o modifican sus conductas.

Particularmente, las conductas alimentarias son de carácter complejo y de origen multifactorial y condicionan directamente el estado nutricional de la persona (Troncoso, 2009). Éstas se encuentran asociadas a las preferencias y diversos estudios han tratado de precisar los determinantes que influyen en las elecciones. Respecto a los individuos adultos, los factores que inciden con mayor relevancia son el sabor, el costo, la nutrición, la conveniencia, el placer y cuestiones éticas y religiosas, entre otras (Glanz *et al.*, 1998; Lindeman & Väänänen, 2000; Steptoe *et al.*, 1995).

Dichos factores, pueden estar presentes en las conductas alimentarias de los jóvenes aunque no necesariamente todos y con la misma ponderación. Conocer más sobre los aspectos involucrados en sus elecciones alimentarias permitiría identificar posibles sesgos de un comportamiento racional. Las preguntas de investigación planteadas son *¿cómo toman sus decisiones alimentarias los jóvenes universitarios?, ¿qué factores inciden en las mismas?* Es

---

<sup>4</sup>Este trabajo se basa en el informe final de la Beca de Estudiante Avanzado (UNMdP) de Luciano Menéndez, bajo la dirección de la Lic. (Dra.) Miriam Berges (UNMdP) y la codirección de la Lic. Beatriz Lupín (UNMdP), 2019-2020.

por ello que el objetivo del presente trabajo es explorar la incidencia de factores vinculados a las preferencias y a las percepciones de los estudiantes de la educación superior respecto al grado de independencia con que toman sus decisiones alimentarias.

Se sigue el eje conceptual de la Economía del Comportamiento que estudia limitaciones a la racionalidad en la conducta de los individuos. Lo anterior, supone inconsistencias cuando se deben tomar decisiones de largo plazo ya que costos y beneficios ocurren en distintos momentos (Thorgeirsson & Kawachi, 2013).

## MATERIALES Y MÉTODOS

La población objetivo del estudio es la comunidad estudiantil de la Facultad de Ciencias Económicas y Sociales-Universidad Nacional de Mar del Plata (FCEyS-UNMdP). Se elaboró un instrumento de recolección de datos que 229 estudiantes completaron, guiados por el primer autor de este Trabajo, durante los meses de abril y junio del año 2019. Dicho instrumento, estaba estructurado en 5 ejes: características y hábitos alimentarios, antecedentes de salud, conocimientos nutricionales, indicadores subjetivos sobre la salud y datos demográficos y socioeconómicos. El relevamiento se realizó en el ámbito áulico, en horarios cedidos por los docentes responsables de distintas asignaturas, cubriendo las diferentes carreras y años, conforme cuotas de sexo y edad de los registros institucionales. Dado el propósito y alcance de este trabajo y la naturaleza de las variables involucradas, se aplicó un Análisis de Correspondencias Múltiples (ACM) (Johnson & Wichern, 1998; Peña 2002), empleando el *Software* InfoStat® –versión estudiantil–.

## RESULTADOS Y DISCUSIÓN

El grado relativo de independencia en la toma de decisiones actuó como variable guía en el análisis. La misma se construyó a partir de una serie de preguntas del cuestionario referidas a la edad (+/- 25 años) del encuestado, a si convivía con sus padres y a su influencia respecto a los alimentos consumidos en su hogar. Por otra parte, se incluyeron variables que reunían atributos tenidos en cuenta al momento de elegir los alimentos: poder comprarlos cerca del lugar de residencia o de trabajo y que demanden poco tiempo de preparación (Conveniencia), que su marca sea conocida y/o su apariencia y envase sea agradable (Apariencia), que no sean caros y/o tengan buena calidad-precio (Económico), que sean nutritivos y saludables y que satisfagan algún interés particular<sup>5</sup>. También se tomaron los casos donde el sabor y/o aroma de los alimentos sea agradable y el hecho de que los encuestados consideren que muchas opciones saludables no cumplen con tal condición (Gustos). Finalmente, se incluyeron 2 variables relativas a la percepción de los encuestados con respecto a su dieta y los efectos de la misma sobre su salud. La primera de ellas distingue entre los encuestados que consideran su dieta como saludable y aquellos que no<sup>6</sup>. La restante pretende aproximar el grado de susceptibilidad del individuo, diferenciando aquellos que creen muy probable un deterioro de su salud si no adopta una alimentación saludable<sup>7</sup>.

Del ACM, surge que el patrón de asociación que se evidencia más claramente (2°-3° cuadrante, Gráfico 1) se relaciona a los estudiantes con mayor independencia relativa (*Mas Indep*; 79,5%

<sup>5</sup>Incluye considerar el impacto ambiental de la producción de los alimentos, su religión o estilo de vida, que le recuerden orígenes o tradiciones familiares y/o que no contengan aditivos, conservantes y/o colorantes.

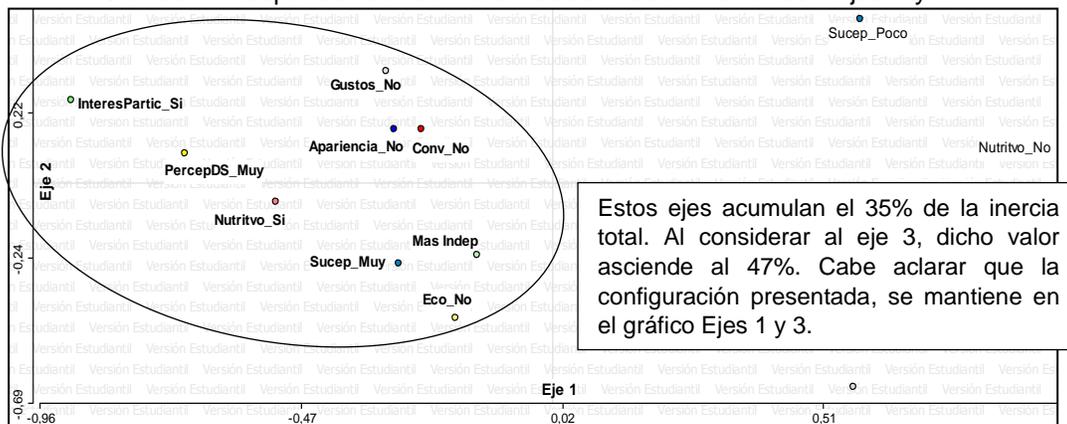
<sup>6</sup>Los encuestados debían señalar qué tan saludable consideraba que era su dieta (escala 1: no saludables a 5: totalmente saludables). Se diferenció entre quienes perciben que su dieta es saludable (4-5) y los que no (1-3).

<sup>7</sup>Los encuestados evaluaron la posibilidad de que su salud se afecte si no adopta una alimentación saludable (escala de 1: nada probable, a 5: totalmente cierto). Los casos se segmentaron entre los más susceptibles (4-5) y los menos susceptibles (1-3).

de la muestra). Aunque los mismos perciben su dieta como saludable (*PercepDS\_Muy*) y se consideran susceptibles a problemas de salud en caso de no mantener una alimentación adecuada (*Sucep\_Muy*; 70% de la encuesta), también, admiten que ciertos hábitos no saludables, son difíciles de cambiar. De todos modos, la mayor independencia para tomar decisiones alimentarias parecería vinculada a una mayor conciencia sobre el impacto de la alimentación en la salud. Estos encuestados señalan como relevante los alimentos sean nutritivos (*Nutritivo\_Si*) y que se adapten a sus intereses particulares (*InteresPartic\_Si*). El primero de ellos es considerado en la mayoría de los casos (67%) pero solo el 25% manifiesta adecuar sus elecciones a algún interés. En relación a esto último, el 19% de los encuestados indica seguir alguna dieta. En cambio, no resultarían determinantes para estos estudiantes la conveniencia (*Conv\_No*), la apariencia (*Apariencia\_No*), lo económico (*Eco\_No*) y los Gustos (*Gustos\_No*): 72%, 74%, 65% y 66, respectivamente.

Los resultados sugieren que la mayor responsabilidad que conlleva su autonomía se traduce en un mayor grado de conciencia sobre el vínculo alimentación-salud, actuando sobre los factores que influyen en las decisiones. Esto podría dar algún indicio respecto a la consistencia intertemporal de sus decisiones y, con ello, sobre las limitaciones de su fuerza de voluntad. Sin embargo, resulta difícil suponer que los aspectos relativos a gustos y a conveniencia no sean relevantes. Se prevé seguir avanzando en el estudio, extendiéndolo a otras facultades y ampliando el número de casos a fin de precisar resultados mediante la estimación de un modelo econométrico.

Grafico 1: Independencia relativa en decisiones alimentarias –Ejes 1 y 2–



Fuente: elaboración propia en base a la fuente primaria de datos. Software InfoStat® –versión estudiantil–.

## REFERENCIAS

- [1] Glanz, K. *et al.* (1998). Why Americans eat what they do: Taste, nutrition, cost, convenience, and weight control concerns as influences on food consumption. *Journal of the American Dietetic Association*, 98: 1.118-1.126.
- [2] Johnson, R. A. & Wichern, D. W. (1998). *Applied Multivariate Statistical Analysis*. USA: Prentice-Hall.
- [3] Lindeman, M. & Väänänen, M. (2000). Measurement of ethical food choice motives. *Appetite*, 34(1): 55-59.
- [4] Peña, D. (2002). *Análisis de Datos Multivariantes*. Madrid-España: McGraw Hill/Interamericana de España S. A.



- [5] Steptoe, A.; Pollard, T. M. & Wardle, J. (1995). Development of a measure of the motives underlying the selection of food: the food choice questionnaire. *Appetite*, 25(3): 267-284.
- [6] Thorgeirsson, T., & Kawachi, I. (2013). Behavioral economics: merging psychology and economics for lifestyle interventions. *American Journal of preventive medicine*, 44(2), 185-189.
- [7] Troncoso, C. & Amaya, J. (2009). Factores sociales en las conductas alimentarias de estudiantes universitarios. *Revista Chilena de Nutrición*, 36(4): 1.090-1.097.

## CONTAMINACIÓN VEHICULAR. MODELADO GEOESTADÍSTICO CON VARIOGRAMAS DIFUSOS

MOLINARI DANIELA A.<sup>1</sup>, CHAPARRO MAURO A.E.<sup>1</sup>, CHAPARRO MARCOS A.E.<sup>2</sup>

<sup>1</sup>Centro Marplatense de Investigaciones Matemáticas (CEMIM-UNMDP), CONICET

<sup>2</sup>Centro de Investigaciones en Física e Ingeniería del Centro de la Provincia de Buenos Aires (CIFICEN-UNCPBA), CONICET  
daniela.a.molinari@gmail.com

### RESUMEN

En esta contribución, se propone una alternativa para el análisis de una variable difusa en un estudio geoestadístico. Con el objetivo de realizar una predicción espacial de la contaminación vehicular en la ciudad de Mar del Plata e identificar zonas potencialmente contaminadas, se combinaron elementos de geoestadística y matemática difusa. La variable en estudio es el Índice Magnético de Contaminación (IMC), un sistema de inferencia difusa, cuya respuesta es un número difuso. La zona de estudio está formada por 780 manzanas del micro y macrocentro de la ciudad de Mar del Plata de las que se muestrearon 46 puntos. La dependencia espacial se estimó por medio de un variograma empírico difuso, utilizando aritmética difusa para su cálculo y, se ajustó visualmente un modelo teórico exponencial. A partir del mismo, se realizó un mapa de predicción por medio de un Kriging ordinario para diferentes  $\alpha$ -cortes. La aplicación de esta metodología permitió realizar predicciones en zonas que no habían sido muestreadas, con el agregado de poder cuantificar la incertidumbre de la medición en cada punto predicho con un número difuso. A partir de los mapas de predicción se determinaron tres zonas con niveles de contaminación superiores a zonas denominadas “limpias”.

**Palabras clave:** *número difuso, aritmética difusa, kriging ordinario, magnetismo ambiental, índice magnético de contaminación.*

### INTRODUCCIÓN

El monitoreo y la evaluación de la calidad del aire que se realiza a través de la utilización de técnicas de magnetismo ambiental, comprende una descripción sobre la concentración, mineralogía y el tamaño de grano magnético. Así, la variable que describe el estado del ambiente es de naturaleza multivariada. Para reducir la dimensión del problema, en este trabajo, se consideró como variable respuesta el *Índice Magnético de Contaminación* (IMC); un sistema de inferencia difusa de tipo Mamdani construido por Chaparro *et al.* (2015a) ([3]) de respuesta difusa.

Para el tratamiento de la información imprecisa del IMC, caracterizada en el sentido de la teoría de conjuntos difusos; utilizaremos la aritmética de números difusos. Cuando la variable en estudio es un número difuso el modelado de la correlación y la predicción espacial deben ser adaptados. El modelado de la correlación se realiza a través de la construcción de variogramas empíricos difusos y se ajustan modelos teóricos para diferentes  $\alpha$ -cortes. Por otra parte, en la predicción se utiliza el método Kriging que tiene la particularidad de incorporar a la misma la función de variograma construida previamente.

El objetivo de este trabajo es realizar una predicción del área de estudio para evaluar zonas con niveles de contaminación vehicular en la ciudad de Mar del Plata. Además, aportar una

alternativa de análisis cuando la variable toma valores difusos, en nuestro caso, números difusos triangulares.

## MATERIALES Y MÉTODOS

A fines del mes de marzo y principios de abril del año 2017 se recolectaron 46 muestras de corteza de árboles distribuidas en un grillado regular de 8 km<sup>2</sup>. El mismo abarcó parte de la zona costera, microcentro, uno de los macrocentros comerciales y zonas residenciales de la ciudad de Mar del Plata. La distancia entre cada sitio de muestreo fue de 400 m en dirección Norte-Sur, y 300 en dirección Este-Oeste. En cada sitio de muestreo, se seleccionó un árbol y se tomó una muestra de su corteza a una altura de 0.5 m. En el laboratorio del Grupo de Magnetismo Ambiental (CIFICEN, CONICET-UNCPBA) en Tandil, se prepararon y midieron las muestras, obteniéndose, entre otros, los cuatro parámetros  $\chi$ , MRIS,  $H_{cr}$ , y  $\chi_{MRA}/\chi$ -ratio, necesarios para el cálculo del IMC a través del sistema de inferencia difusa.

En este trabajo, se aplica aritmética difusa mediante la descomposición por sus  $\alpha$ -cortes que permite trabajar con aritmética de intervalos cerrados (ver [4]). Un *subconjunto difuso*  $\tilde{A}$  de  $X$  es dado por un conjunto de pares ordenados,  $\tilde{A} = \{(x, \mu_{\tilde{A}}(x)), \text{ con } x \in X\}$  donde  $\mu_{\tilde{A}}(x)$  es la función de pertenencia de  $x$  en  $\tilde{A}$ .

El conjunto de todos los elementos que pertenecen al menos con grado  $\alpha$  al conjunto difuso  $\tilde{A}$ , es llamado conjunto  $\alpha$ -corte, es decir,  $\tilde{A}_\alpha = \{x \in X / \mu_{\tilde{A}}(x) \geq \alpha\}$  para  $0 \leq \alpha \leq 1$ .

Entonces, se considera un *número difuso triangular*  $\tilde{A} = (a, b, c)$  a un conjunto difuso cuya función de pertenencia es:

$$\mu_{\tilde{A}}(x) = \begin{cases} 0 & \text{si } x < a \\ \frac{x-a}{b-a} & \text{si } a \leq x \leq b \\ \frac{c-x}{c-b} & \text{si } b \leq x \leq c \\ 0 & \text{si } x > c \end{cases}$$

con  $a < b < c$  números reales.

Así definidos, es posible realizar las operaciones (suma, multiplicación, multiplicación por escalar y potencia) con aritmética de intervalos de sus  $\alpha$ -cortes.

El uso de datos imprecisos es esencial en el proceso de definir la estructura espacial de una variable regionalizada. Este tipo de datos se proporcionan en forma de números difusos, por lo que los valores resultantes del variograma empírico también serán números difusos. Recordemos, la expresión clásica del variograma empírico:

$$\hat{\gamma}(h) = \frac{1}{2|N(h)|} \sum_{i=1}^{N(h)} (Z(s_i) - Z(s_i + h))^2 \quad (1)$$

donde  $|N(h)|$  es el número de pares de lugares de muestreo situados a una distancia  $h$ .

Siguiendo la propuesta de Bárdossy *et al.* 1990 ([2]) se calcula el variograma empírico difuso. Una vez calculado el variograma empírico se debe ajustar de forma visual un variograma teórico, para una cantidad finita  $\alpha$ -cortes (en este caso,  $\alpha=0,0.5,1$ ), ya que se considera que en el caso de variogramas difusos este ajuste es más apropiado.

Una vez ajustados los variogramas teóricos para cada uno de los  $\alpha$ -cortes, estos se utilizan aplicándolos sobre un Kriging ordinario. De esta manera, las estimaciones de cada punto y para cada  $\alpha$ -corte se presentan como números difusos. El ancho de la estimación se interpreta como una medida de incertidumbre que depende de los datos.

## RESULTADOS Y DISCUSIÓN

La teoría de conjuntos difusos, es una herramienta útil, realista y relativamente sencilla a la hora de modelar la imprecisión. El desarrollo de la aritmética difusa y, principalmente, las operaciones por medio de  $\alpha$ -cortes, constituye una forma computacionalmente práctica de ser implementada. La aplicación de esta metodología permitió realizar predicciones en zonas que no habían sido muestreadas con el agregado de poder cuantificar la incertidumbre de la medición en cada punto predicho (Fig. 1). Este tipo de predicciones si bien carece de significancia estadística permite a los investigadores tener una primera “vista” de la problemática y establecer nuevos (o reafirmar) criterios como, por ejemplo, en el diseño de muestreo de nuevas campañas. A partir de la lectura de los mapas de predicción, se observaron tres focos de contaminación, el valor de  $IMC > 2$ :

- En la zona cercana a las calles Sarmiento y Alberti, donde hay un tránsito intenso y a baja velocidad, se observaron valores de IMC de máxima pertenencia comprendidos entre 2 y 3 con una amplitud (o incertidumbre) que va desde 1 a 3.
- Un foco más amplio en la zona cercana a la Avenida Independencia y calle Falucho con valores de IMC comprendidos entre 2.5 y 3.5 de máxima pertenencia (amplitud de 1 a 5).
- En la zona de las calles San Juan y Necochea se presentaron los mayores valores de IMC, entre 2.5 y 4 (amplitud de 3 a 5).

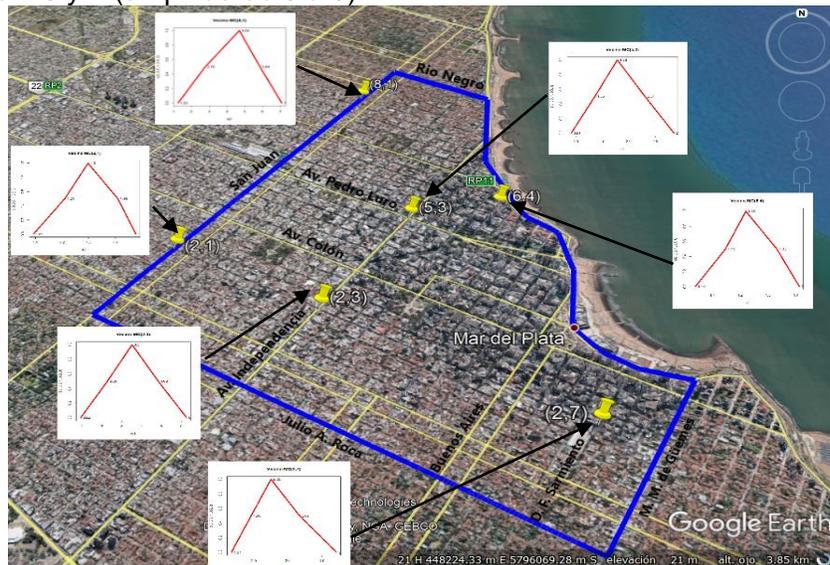


Fig. 1. Valores predichos para un vecino de cada uno de los seis sitios muestreados seleccionados en la zona de estudio.

## REFERENCIAS

- [1] Banerjee, S. Carlin, B.P., Gelfand A.E. (2004). *Hierarchical Modeling and Analysis for Spatial Data*. Chapman & Hall/CRC, Estados Unidos.
- [2] Bárdossy A., Bogardi I., Kelly W.E. (1990). “Kriging with imprecise (fuzzy) variograms. I: Theory”. *Mathematical Geology*, 22(1): 63-79.
- [3] Chaparro M.A.E., Chaparro M.A.E., Marié D.C., Casteñeda-Miranda A.G., Sinito A.M., Böhnel H.N. (2015a). “Estimación de un Índice Magnético de Contaminación a través de un sistema de inferencia difusa”. *Geos*, 35(1): 81.



[4] Hanss, M. (2005). *Applied Fuzzy Arithmetic. An Introduction with Engineering Applications*. Springer, Alemania.

## RENDIMIENTO ACADÉMICO EN ÓPTICA GEOMÉTRICA

MONTENEGRO BRUSOTTI JUDITH<sup>1,2</sup>, PONTEVILLE CHRISTIANE<sup>1</sup>, NÚÑEZ MYRIAM<sup>1</sup>

<sup>1</sup>Cátedra de Matemática, Facultad de Farmacia y Bioquímica, Universidad de Buenos Aires

<sup>2</sup>Cátedra de Óptica, Facultad de Farmacia y Bioquímica, Universidad de Buenos Aires

*jmontenegro@ffyb.uba.ar*

### RESUMEN

El objetivo de este trabajo es realizar un análisis cuantitativo del rendimiento académico de los alumnos de la asignatura Óptica Geométrica de la Tecnicatura Universitaria en Óptica y Contactología, durante el primer cuatrimestre de 2020 en la Facultad de Farmacia y Bioquímica de la U.B.A. Este análisis busca determinar si la evaluación inicial proporciona una herramienta que nos permita rever estrategias de enseñanza que mejoren el rendimiento académico posterior de los alumnos.

Este trabajo concibe al rendimiento académico como el análisis de la historia académica de los alumnos a lo largo de las diferentes instancias que forman parte de la evaluación de la asignatura. Algunas de las variables consideradas fueron: número de ejercicios de la prueba diagnóstica (PD) clasificados por categorías ( $G_1$ : Matemática,  $G_2$ : Física,  $G_3$ : Geometría), número de parciales y recuperatorios rendidos y aprobados y condición final de cursada.

Se efectuó una partición de los resultados de la PD en tres grupos, considerando un puntaje de ejercicios bien resueltos (Primer grupo: menor o igual a 2, Segundo Grupo: igual a 3, Tercer grupo: entre 4 y 5).

Se observó que, el porcentaje de alumnos que regularizaron la asignatura para el primer grupo fue, aproximadamente, del 67,6%. Para los estudiantes del segundo grupo, el porcentaje que regularizó la asignatura fue de 29,7%, mientras que para los del tercer grupo el porcentaje de alumnos que regularizó la materia fue del 2,7%.

Los temas en los que los alumnos presentaron mayor dificultad corresponden a los del grupo  $G_2$ . Se comprobó que no hay independencia entre la aprobación de la prueba diagnóstica y la condición al final del cuatrimestre  $p < 0,0001$ .

Se identificó una importante mejora académica de los alumnos que desaprobaban la prueba diagnóstica y regularizaron la asignatura.

**Palabras clave:** *rendimiento académico, independencia, estrategias de enseñanza en matemática.*

### INTRODUCCIÓN

En la actualidad, el análisis cuantitativo del rendimiento académico constituye uno de los elementos utilizados para caracterizar la historia académica de los alumnos, a lo largo de su trayecto institucional, en los diferentes niveles de la educación. En este marco, se pueden analizar las diferentes instancias curriculares y la evolución personal y grupal a lo largo de la asignatura.

Uno de los objetivos de este trabajo es determinar si la prueba diagnóstica inicial proporciona herramientas para explicar los recorridos académicos y, de esta forma, nos permita rever estrategias de enseñanza que mejoren el proceso de enseñanza y aprendizaje de los alumnos.

El otro objetivo es comparar los resultados obtenidos en las evaluaciones y condición final de los estudiantes que cursaron la asignatura anteriormente mencionada en los años 2019 y 2020,

como también el análisis de las diferencias entre la evaluación presencial y la virtual en el contexto actual de pandemia.

## **MATERIALES Y MÉTODOS**

Se analizó el recorrido y evolución académica de un total de 108 alumnos cursantes de Óptica Geométrica en 2020. La prueba diagnóstica se realiza el primer día de clase virtual, con el fin de evaluar los saberes previos necesarios para la cursada de la asignatura. Se efectuó una partición de los resultados de la prueba diagnóstica en tres grupos (Primer grupo: menor o igual a 2, Segundo Grupo: igual a 3, Tercer grupo: entre 4 y 5). Algunas de las variables consideradas fueron: cuatrimestre de cursada, año de ingreso a la unidad académica, número de ejercicios de la prueba diagnóstica clasificados por temáticas ( $G_1$ : Matemática,  $G_2$ : Física,  $G_3$ : Geometría), número de parciales aprobados/desaprobados, número de recuperatorios rendidos, número de recuperatorios aprobados, condición final de cursada.

Se realizó un análisis descriptivo de los datos. El análisis de la información obtenida se realizó mediante la Prueba de Chi-Cuadrado.

## **RESULTADOS Y DISCUSIÓN**

El primer análisis corresponde al de los estudiantes que cursaron la asignatura durante el Primer Período Lectivo 2020.

Del total, el 42,7% alcanzó la aprobación de las actividades obligatorias de la asignatura (“regularizó”).

Los temas en los que los alumnos presentan mayor dificultad corresponden a la temática referente al grupo  $G_2$ .

Se comprobó que no hay independencia entre la aprobación de la prueba diagnóstica y la condición al final del cuatrimestre ( $p < 0,0001$ ; Prueba de Chi cuadrado), ni entre las variables: condición final y año de cursada ( $p < 0,0001$ ; Prueba de Chi cuadrado). El número de aprobados bajo la modalidad virtual fue levemente menor.

Asimismo, aquellos que demostraron mejores resultados en la evaluación diagnóstica, alcanzaron un mayor número de exámenes regulatorios aprobados.

En el análisis de las diferencias entre los rendimientos académicos de los años 2019 y 2020, se observó que, en el año 2020, el porcentaje de alumnos que regularizaron la asignatura para el primer grupo fue, aproximadamente, del 42,7% y del 48% en el 2019.

Se identificó una importante mejora académica de los alumnos que desaprobaron la prueba diagnóstica y regularizaron la asignatura.

En cuanto al análisis vinculado con los ejes temáticos de la prueba diagnóstica se evidenciaron mayores dificultades en los temas vinculados a Física.

## **REFERENCIAS**

[1] Nuñez, M.; Granchetti, H.; Ponteville, CH. (2017). “Evaluación del rendimiento académico en bioestadística”. *Acta Latinoamericana de Matemática Educativa*, 30(1): 782-789.

[2] Nuñez, M.; Fericola, M.; Castro, M.J.; Ponteville, CH. (2017). “Análisis estadístico del nivel de Aprendizaje en Matemática y Estadística”. *Acta Latinoamericana de Matemática Educativa*, 30(1): 2448-6469.

[3] Nuñez, M.; Bonadías, M.E.; Montenegro Brusotti, J.; Ponteville, CH. (2019). “Evolución Académica en Matemática Aplicada”. XXIV Reunión Científica del Grupo Argentino de Biometría. Mendoza. República Argentina.



[4] Nuñez, M.; Bonadías, M.E.; Montenegro Brusotti, J.; Ponteville, CH. (2018). “Rendimiento Académico en Matemática”. XXIII Reunión Científica del Grupo Argentino de Biometría. II Encuentro Argentino-Chileno de Biometría. Neuquén. República Argentina.

## ALTERNATIVAS PARA MEDIR LA VARIABILIDAD ESPACIAL DE LA MATERIA ORGÁNICA DEL SUELO A PARTIR DE SENSORES VERSUS MUESTREOS DE LABORATORIO

MUÑOZ H. M.<sup>1</sup>, ZABALA S. M.<sup>1</sup>, VARGAS S. L.<sup>1</sup>, FILIPPINI O. S.<sup>1</sup>, VÁZQUEZ J. M.<sup>2</sup>

<sup>1</sup>Departamento de Ciencias Básicas, Universidad Nacional de Luján

<sup>2</sup>Departamento de Tecnología, Universidad Nacional de Luján

mhm\_1602@hotmail.com

### RESUMEN

El incremento del carbono orgánico (CO) del suelo permite aumentar la productividad en forma sustentable. Sin embargo, medir las variaciones en el stock de carbono de los suelos es dificultoso y solo son apreciables luego de varios años, dificultando tomar decisiones. Existen diferentes alternativas de muestreo y de uso de distintas fracciones lábiles del CO del suelo para medir en forma temprana sus cambios. Sin embargo, no existen estudios que evalúen integradamente la variabilidad espacial del CO. Recientemente, fue introducida la detección óptica en dispositivos montados en sembradoras agrícolas. Mediciones precisas de estos dispositivos podrían proporcionar mapeos escalables de materia orgánica (MO) de bajo costo. El objetivo de investigación fue evaluar el desempeño de estos dispositivos versus resultados de laboratorio y obtener una representación euclídea respetando las relaciones de proximidad entre las características medidas: MO, humedad, además de las relaciones entre estos parámetros edáficos y otros como rendimiento de cultivos obtenidos a partir de monitores de flujo de grano en cosechadoras. En el mismo lote fue realizado un muestreo de suelo a partir de una grilla, con celdas de una hectárea, según un semivariograma. El contenido de MO de las muestras fue obtenido por oxidación húmeda, en laboratorio. Utilizando Modelos de Regresión Lineal fueron estudiadas las relaciones entre el MO y características como rendimientos, elevación, flujo de cultivo, humedad y otros parámetros. Fueron relacionados los datos de MO de los sensores versus laboratorio, que produjo una correlación del 35%. El escalamiento multidimensional de datos de disimilaridad, procedente de estas múltiples fuentes permitió encontrar la estructura de un conjunto de medidas de distancia entre elementos dentro de un espacio bidimensional. Las dimensiones de este espacio conceptual fueron interpretadas y mejoraron la comprensión de los datos. La bondad del ajuste se representó por el Stress de 0,002 y  $R^2$  de 0,999.

**Palabras clave:** *variabilidad espacial, sensores, escalamiento multidimensional, egresiones.*

### INTRODUCCIÓN

El uso de sensores de reflectancia para la estimación de la MO del suelo directamente en el surco de siembra o los rindes a cosecha, ha sido utilizado, a nivel global desde hace una década y en los últimos años, en Argentina. Estos posibilitaron la confección de mapas de MO del suelo a una escala adecuada para realizar un manejo sitio específico de la fertilidad por lo que un objetivo de este trabajo fue evaluar su desempeño. Conway *et al.* (2019), evaluaron la capacidad de predecir el contenido de MO del suelo por reflectancia dentro de un surco de siembra mediante el sensor SmartFirmer, que es el utilizado en este estudio. Para contrastar

con los datos de laboratorio, fueron promediados los datos del sensor (sembradora) de reflectancia generados dentro de los 5 metros de los puntos de muestreo. Fue evaluado el desempeño de estos dispositivos versus resultados de laboratorio y fue obtenida una representación euclídea que respetó las relaciones de proximidad entre las características medidas.

## **MATERIALES Y MÉTODOS**

El estudio fue realizado en un predio de la localidad de Carmen de Areco ( $34^{\circ}26'S$ ,  $59^{\circ}44'O$ ), en un suelo Argiudol Acuico (*Soil Taxonomy*, 1975), de la serie Gouin, en un paisaje de planicie suavemente ondulada en posición de lomas bajas y planos levemente deprimidos de la Subregión Pampa Ondulada alta. Los valores de MO estimados en cada sensor surgieron de un promedio de sus lecturas de reflectancia y las de los sensores adyacentes. Además, mapas de rendimiento de los cultivos generados por el sensor de rendimiento de una cosechadora de granos fueron obtenidos.

Con posterioridad a la emergencia del cultivo fue realizado un muestreo en grilla tomando una muestra compuesta por una hectárea. Para definir las dimensiones de las celdas de la grilla fue adoptado el criterio de que la distancia entre las muestras sea igual a un tercio del rango de un semivariograma, generado a partir de información previa sobre la variabilidad espacial del lote como imágenes satelitales o mapas de rendimiento (Oliver, M.A. 2010), aunque en este caso fue tomado como referencia el semivariograma generado con los datos de MO obtenidos con el sensor de reflectancia de la sembradora. En cada punto de muestreo fue tomada una muestra compuesta por 5 submuestras, una en el centro y las otras cuatro a dos metros del punto central formando una cruz. Cada submuestra de suelo fue tomada con barreno hasta los 5 cm para coincidir con la profundidad evaluada por los sensores de reflectancia. Para que la localización de los puntos en el terreno sea precisa, fue utilizado un receptor GNSS de doble frecuencia. Las muestras de suelo fueron secadas al aire y molidas hasta que puedan pasar por un tamiz de 500 micrones. El contenido de MO de las muestras fue determinado por el método de (Walkley & Black, 1934), bajo la Norma IRAM-SAGyP 29571-3 (2011). Para poder comparar las dos metodologías fueron calculados los valores medios de los datos de MO estimados por los sensores de reflectancia, para las coordenadas de los puntos donde fueron tomadas las muestras de suelo con barreno. Las medias fueron calculadas con el software QGIS 3. Con las medias MO suelo estimado por reflectancia y los medidos en el laboratorio, fue analizada la relación entre los valores obtenidos por los distintos métodos a partir de un Análisis de Regresión Múltiple, considerando la bondad del ajuste y el coeficiente de determinación  $R^2$ . Posteriormente, fue utilizado análisis multivariante, específicamente Escalamiento Multidimensional, partiendo de una matriz de distancias entre individuos a partir de los datos de las diversas fuentes, sensores y laboratorio, y fue producida una representación de los individuos en una escala euclídea ordinaria de modo que las distancias en dicha escala se aproximen lo mejor posible a las distancias de partida. El procedimiento de escalamiento utilizado (Alscal) fue elegido, porque es el único que permite realizar análisis métrico. Fue buscada la convergencia global del proceso iterativamente. El criterio mínimo cuadrático de optimización denominado SSTRESS y  $R^2$ , fueron calculados.

## **RESULTADOS Y DISCUSIÓN**

Las zonas identificadas por el sensor de la cosechadora como A (Fig.1) presentaron una relación positiva entre el contenido de MO y el rendimiento. Las zonas identificadas como B, a pesar de tener altos contenidos de MO, presentaron bajos rendimientos por situarse en

posiciones deprimidas en el paisaje, donde ocurrieron condiciones de anegamiento temporal en un año húmedo. Entre MO del suelo y los rendimientos, hubo una correlación positiva baja del 13 % y significativa del 40 % entre rindes y MO de laboratorio.

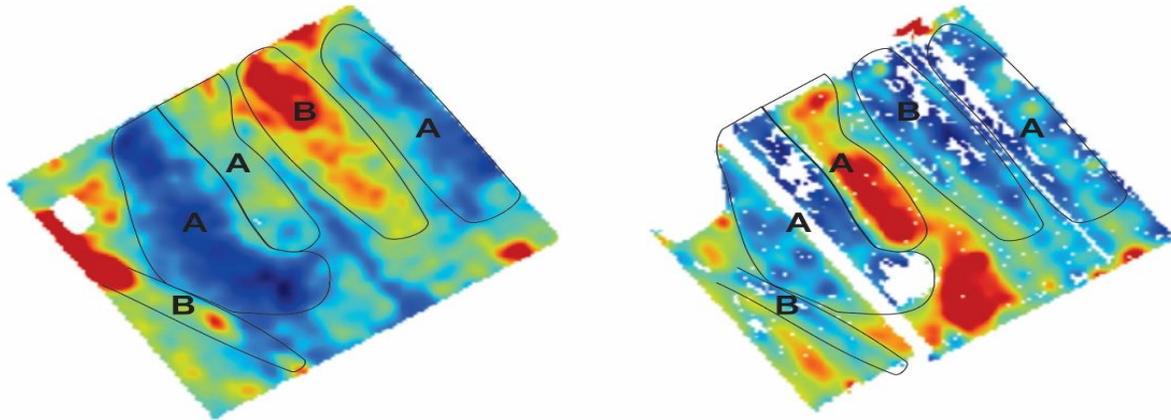
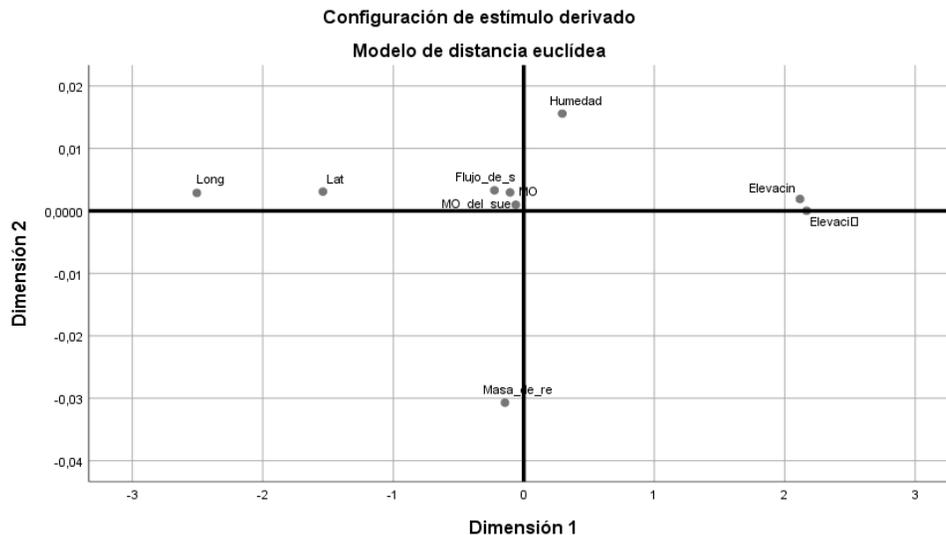


Figura 1. Mapas de materia orgánica del suelo (izq.) y rendimiento de soja (der.).

El relevamiento de la variabilidad espacial de la MO del suelo mediante sensores de reflectancia midiendo dentro de los surcos de siembra presentó una relación positiva pero débil con la medición del mismo parámetro por oxidación húmeda de muestras de suelo obtenidas a partir de grillas de muestreo en laboratorio del 35 %.



Fue obtenida una representación euclídea que respeta lo más posible las relaciones de proximidad entre las determinaciones de Mo del suelo a laboratorio y MO determinada por los sensores, sobre las mismas unidades experimentales. Fue medida la bondad de ajuste entre las distancias de la configuración y las disparidades, Stress de 0.0002, RSQ de 0,999. Para obtener mapas de la distribución de la MO fue necesario realizar ajustes, como desarrollar algoritmos de calibración de los sensores.



## **REFERENCIAS**

- [1] Conway, Lance S., et al. (2019). Planting Depth and Soil Series Effect on In-Furrow Soil Sensor Performance. En Sudduth, Kenneth A., Kitchen, Newell R., & Veum, K.S. (Eds.), 5th Global Workshop on Proximal Soil Sensing Proceedings. [www.pss2019.org](http://www.pss2019.org)
- [2] Borg, I; Groenen, P.; Mair, P.; (2013), *Applied Multidimensional Scaling* Springer Heidelberg New York Dordrecht London, pág.1-113.
- Husson, F. & Le, S. (2007). Sensominer: Sensory data analysis with R (R package version 1.07).
- [3] Oliver, M.A. 2010. Geoestatistical Applications for Precision Agriculture. Springer Science & Business Media.

## **INVESTIGACIÓN DEL HÁBITO TABÁQUICO EN PACIENTES ONCOLÓGICOS, EN ENFERMEDADES NEOPLÁSICAS CON Y SIN RELACIÓN CON EL CONSUMO DE CIGARRILLOS**

NUÑEZ M.<sup>1</sup>, BONADIES E.<sup>1</sup>, VERZURA M.A.<sup>2</sup>, MAUR PEROTTI J.A.<sup>2</sup>, PEREZ MESA P.<sup>2</sup>,  
SPECTERMAN S.<sup>2</sup>, LASTIRI J.M.<sup>2</sup>

1. *Cátedra de Matemática. Facultad de Farmacia y Bioquímica. UBA*

2. *Hospital Italiano de Buenos Aires*

*myriam@ffyb.uba.ar*

### **RESUMEN**

El tabaquismo constituye la principal causa de muerte evitable en los países desarrollados. Los cinco principales riesgos para la salud están relacionados con el comportamiento y la alimentación: índice de masa corporal alto, bajo consumo de frutas y hortalizas, inactividad física, tabaquismo e ingesta excesiva de alcohol. Estos factores causan el 30% de las muertes por cáncer. El consumo de tabaco es el más importante: causa 22% de las muertes mundiales por cáncer en general y 71% de las muertes por cáncer de pulmón.

Los objetivos de este trabajo son: a) conocer la prevalencia y características del tabaquismo, y la conciencia de daño en pacientes con cáncer a través de una encuesta realizada en el período 2016-2017 (Grupo Nuevo: GN). Comparar las características del hábito en pacientes con neoplasias asociadas y no asociadas. Comparar estos resultados con la población estudiada con la misma encuesta realizada en el período 1997-1998 (Grupo Anterior: GA), y b) Conocer la repercusión que han tenido las nuevas leyes de protección contra el tabaco en la conciencia y prevalencia del hábito.

Ambas encuestas se realizaron en el Hospital Italiano de Buenos Aires (HIBA). Se obtuvieron muestras de 364 y 350 pacientes, respectivamente. Se registraron los siguientes datos: Sexo, Edad, Factores de riesgo: hábito de fumar, frecuencia de cigarrillos, tiempo que lleva fumando, Características del hábito, Conocimiento acerca de la enfermedad, Deseos de cesación, Conciencia de daño.

La aplicación de metodología estadística multivariada como, Análisis Factorial de Correspondencias, Regresión Logística y Árboles de Clasificación, nos permitió tener una idea del comportamiento de todos los factores de riesgo en su conjunto.

Si bien se hallaron diferencias entre los períodos estudiados que hablan del beneficio que produjeron las nuevas leyes de protección, consideramos que es fundamental seguir trabajando para que estas diferencias sean más notorias.

**Palabras clave:** *hábito tabáquico, análisis factorial de correspondencias, regresión logística.*

### **INTRODUCCIÓN**

Las enfermedades no transmisibles (ENT) son actualmente la principal causa de mortalidad mundial, principalmente, enfermedades cardiovasculares, cáncer, diabetes y enfermedades pulmonares crónicas. Un alto porcentaje de ENT son prevenibles y comparten los mismos factores de riesgo. Aunque la edad es un factor de riesgo importante, el consumo de tabaco es el más importante: causa 22% de las muertes mundiales por cáncer en general y 71% de las

muertes por cáncer de pulmón. El 40% de los cánceres podrían evitarse con acciones sencillas como no fumar, hacer ejercicio regularmente y mantener una dieta saludable. La OMS remarca que el tabaquismo es la principal causa evitable de cáncer en el mundo.

Las variables asociadas con el desarrollo y perpetuación del tabaquismo incluyen la precocidad en el inicio del tabaquismo, el consumo en el hogar o en grupo familiar, entre otras.

En los últimos años en nuestro país se han desarrollado y aprobado varias leyes que intentan frenar este hábito y concientizar a la población.

Consideramos importante conocer el impacto de estas nuevas leyes para comprender si tanto estas nuevas regulaciones, como las campañas realizadas, resultaron eficaces en nuestra población.

## **MATERIALES Y MÉTODOS**

Ambas encuestas se realizaron en el HIBA en los períodos de agosto 2016 – agosto 2017, y marzo 1997 – marzo 1998. Se obtuvieron muestras de 364 pacientes y 350, respectivamente. Se registraron datos acerca de: Cooperación con el encuestador; Características del hábito; Conocimiento sobre de la enfermedad; Datos sobre cesación; Conciencia de daño; Datos personales: Nombre, Apellido, Diagnóstico, Sexo, Edad. El objetivo del estudio es conocer la prevalencia y características del hábito tabáquico, así como la conciencia de daño en pacientes con cáncer. Comparar las características del hábito en pacientes con neoplasias asociadas y no asociadas.

Se realizó un análisis descriptivo para caracterizar la población de cada grupo. Para comparar ambos grupos se aplicaron, de acuerdo con el tipo de variable y su distribución: prueba para diferencias de proporciones, prueba de  $\chi^2$  y prueba no paramétrica U de Mann-Whitney.

Para estudiar el comportamiento de todos los factores de riesgo sobre la incidencia del cáncer de pulmón, se aplicaron los métodos de Regresión Logística, Análisis Factorial de Correspondencias y Árboles de Clasificación.

## **RESULTADOS Y DISCUSIÓN**

La edad media fue: GN:  $62 \pm 14$  años, y GA:  $53 \pm 15$  años. En el GN hubo 51,1% hombres y 48,9% mujeres, mientras que, en el GA hubo 52,3% hombres y 47,7% mujeres.

En el GN el 9,2% fumaba, el 51,8% eran exfumadores y el 39% restante nunca fumó. En el GA fumaba el 21,6%, era exfumador el 30,3% y nunca había fumado el 48,1%. En el GN de los fumadores, el 63,3% eran hombres. En el GA este valor fue 61,3%. En ambos grupos se constató una mayoría masculina fumadora para neoplasias relacionadas y no relacionadas.

La mediana de consumo tabáquico en el GN fue 28 años, mientras que en el GA fue 25 años.

El 28,2% de los pacientes fumaba más de 20 cigarrillos diarios, en lugar de 53,8% como había resultado en el trabajo previo. El consumo de más de 20 cigarrillos por día fue significativamente mayor en hombres. De los que fumaban más de 20 cigarrillos diarios 63,5% tenían neoplasias relacionadas y el 36,5% tenían neoplasias no relacionadas. Estos números en la encuesta previa fueron 65% y 35%, respectivamente. De los fumadores con bajo consumo (<10 cigarrillos diarios) 54,7% son mujeres y 45,3% son hombres. En el trabajo anterior estos resultados fueron 60 y 40%, respectivamente.

El 76,2% de los pacientes con neoplasias no relacionadas estaban incluidos en el grupo de bajo consumo, así como 23,8% con neoplasias relacionadas. En el grupo previo esos valores fueron 34,3 y 18%, respectivamente. La diferencia en el consumo entre neoplasias relacionadas y no relacionadas es estadísticamente significativa ( $p = 0,0005$ ). El 70,7% del GN reconoció que

fumaba en su casa, y el 64,4% en su trabajo. Entre los del GA estos valores fueron 87,2 y 82,3%, respectivamente.

Las patologías oncológicas fueron diversas, siendo las más frecuentes: colorrectal 17,2%, pulmonar 15,9%, pancreático 8,4%, mamario 4,7%, entre otras para el GN. En el GA las neoplasias más frecuentes fueron: onco-hematológicas 20,2%, pulmonar 13,7%, mamario 12,3%, cáncer colorrectal 4,6%, entre otras. El 29,4% de las neoplasias estaban relacionadas con el tabaco, mientras que en el trabajo previo lo estaban 28,3%. El 50% de las neoplasias de los fumadores y el 37,6% de las neoplasias de los exfumadores estaban relacionadas con el consumo, mientras que el 17,7% de las neoplasias de los nunca fumadores se relacionaban con el hábito, siendo esta diferencia estadísticamente significativa ( $p = 0,0003$ ). En el grupo previo estos resultados fueron 40, 40 y 13,8%, respectivamente ( $p = 0,0023$ ).

Cuando se les preguntó a los pacientes acerca de la relación del tabaco con su enfermedad (relacionada o no) el 95,7% contestó correctamente, el 2,15% incorrectamente y el 2,15% desconocía. Se observó una mayor tendencia al error en el grupo fumador. En la encuesta previa se hallaron valores de 46,6, 45,6 y 7,8%, respectivamente.

Se aplicó un Modelo de Regresión Logística donde la variable de respuesta fue el diagnóstico (cáncer de pulmón = 1, otros tipos de neoplasias = 0), las variables que resultaron significativas fueron: Sexo, Ex-Fumador y Fumador.

Se realizó un Análisis Factorial de Correspondencias para establecer la relación entre las variables. Se observó que los factores predominantes corresponden al hábito de fumar y al conocimiento sobre la enfermedad y deben tenerse en cuenta a la hora de planificar medidas de prevención y/o detección temprana del cáncer.

Se encontraron diferencias significativas entre los pacientes de ambas encuestas en relación a los pacientes que fuman más de 20 cigarrillos diarios ( $p = 0,003$ ). En ambas encuestas resultó que los pacientes fumaban más de 20 cigarrillos diarios tenían más neoplasias relacionadas con el tabaco.

Las mujeres consumen en general menos de 10 cigarrillos diarios, aunque en esta última encuesta se vio que están incrementando esa cantidad.

Disminuyó la cantidad de pacientes que fuman en su casa ( $p = 0,0004$ ) y en el trabajo ( $p = 0,0009$ ).

En relación a la técnica de Árboles de Clasificación obtuvimos que el Diagnóstico de Cáncer de Pulmón está asociado con: mayor tiempo de consumo, mayor cantidad de cigarrillos consumidos por día, con menor conciencia sobre la enfermedad y con mayor cantidad de hombres fumadores.

Tanto los fumadores como los exfumadores tienen, en su gran mayoría, neoplasias asociadas con el tabaquismo. En nuestra muestra los fumadores de más de 20 cigarrillos diarios fueron la población más asociada a neoplasias del fumador. Se halló una mayor exposición al humo en el trabajo, en los pacientes no fumadores con neoplasias relacionadas.

El nivel de conocimiento acerca del daño que genera el tabaco ronda en el 90%, demostrando una diferencia significativa con la población antes encuestada. Creemos que este resultado esté asociado a mayor difusión sobre información acerca de los daños que produce el tabaco, así como mayor acceso a información sobre las enfermedades.

Con estos resultados, si bien se hallaron algunas diferencias que hablan del beneficio que produjeron las acciones en la legislación como en la difusión, consideramos que es fundamental seguir trabajando para que estas diferencias sean más notorias.

## **REFERENCIAS**

- [1] Nuñez, M.; Lastiri, J.M.; Pallotta, M.G. (2006). "Técnica estadísticas multivariadas. Prevalencia de factores de riesgo de cáncer". VI Congreso Latinoamericano de Sociedades de Estadística. Rosario, República Argentina.
- [2] Lastriri, J.M.; Nuñez, M.; Pallotta, M.G; Perazzo, D.; Perazzo, F. (2000). *Anales de la Fundación Alberto J. Roemmers*. XIII Ed. (281-297). República Argentina.
- [3] Verzura, A.; Maur Perotti, J.A.; Perez Mesa, P.; Nuñez, M; Lastiri, J.M. (2017). "Investigación del hábito tabáquico y noción del daño, en pacientes oncológicos tratados en el Hospital Italiano de Buenos Aires y allegados, en enfermedades neoplásicas con y sin relación al consumo de cigarrillos". XXIII Congreso Argentino e Internacional de Oncología. Buenos Aires, República Argentina.

## UNA CHECKLIST PARA MAXIMIZAR LA REPRODUCIBILIDAD EN ESTUDIOS DE DIVERSIDAD FUNCIONAL

PALACIO FACUNDO X.<sup>1</sup>

<sup>1</sup>División Zoología Vertebrados, Museo de La Plata, Universidad Nacional de La Plata  
*facundo\_palacio@fcnym.unlp.edu.ar*

### RESUMEN

Desde el siglo XXI, la ciencia atraviesa una crisis de reproducibilidad, de la cual la ecología de comunidades no está exenta. En el caso particular de la diversidad funcional (DF), que estima la diversidad de caracteres de las especies vinculados a funciones ecosistémicas, se carece aún de estándares de reproducibilidad. En este trabajo, propongo una checklist para maximizar la reproducibilidad en estudios de DF. Para esto, organicé los elementos esenciales de los estudios de DF en cuatro grandes temas: (A) construcción de la matriz de comunidades, (B) construcción de la matriz de caracteres funcionales, (C) estimación de métricas de DF y (D) ajuste de modelos y reporte de resultados. Se espera que esta checklist brinde una herramienta para replicar estudios previos, y promueva la adopción de estándares de reproducibilidad en el estudio de la ecología de caracteres funcionales.

**Palabras clave:** *comunidades biológicas, diversidad funcional, reproducibilidad.*

### INTRODUCCIÓN

Desde principios de este siglo, la ciencia atraviesa la denominada crisis de reproducibilidad, en la cual la gran mayoría de los estudios científicos son incapaces de ser replicados bajo los mismos métodos y principios [1]. La ecología de comunidades no se encuentra exenta de esta crisis, y debido a que el estudio de este campo implica muchas veces la medición de variables difíciles de controlar en la naturaleza, es particularmente importante describir en detalle los métodos utilizados [2]. En particular, el estudio de la diversidad funcional (DF), definida como el rango y valor de caracteres funcionales en una comunidad biológica, ha logrado un enorme desarrollo en el entendimiento de patrones y dinámicas comunitarias. Sin embargo, se carece aún de estándares de reproducibilidad en esta área. En este trabajo, propongo una checklist para maximizar la reproducibilidad de estudios de DF.

### MATERIALES Y MÉTODOS

En base a la literatura vinculada a estudios de DF, presento una checklist para estudios de DF, con el objetivo de definir estándares de reproducibilidad que puedan ser utilizados en ecología de comunidades basada en caracteres funcionales. Para esto, compilé una serie de elementos esenciales requeridos para reproducir los resultados de DF y organicé los elementos en cuatro grandes temas: (A) construcción de la matriz de comunidades, (B) construcción de la matriz de caracteres funcionales, (C) estimación de métricas de DF y (D) ajuste de modelos y reporte de resultados.

### RESULTADOS Y DISCUSIÓN

(A) *Construcción de la matriz de comunidades*

El primer paso consiste en detallar el protocolo experimental o de muestreo utilizado, indicando la forma y tamaño de las unidades muestrales o experimentales, el tamaño de muestra, la resolución espacial y temporal del estudio. Con la información de este paso, se construye una matriz  $\mathbf{C}_{n \times s}$  de  $n$  sitios  $\times$   $s$  especies (matriz de comunidades). En este punto, debe indicarse si se utilizaron datos de incidencia (presencia-ausencia), abundancia o biomasa para ponderar el efecto de las especies en las comunidades. Si se corrige la detección imperfecta de las especies mediante modelos multiespecies [3], debe reportarse la información relativa a enfoques Bayesianos (distribuciones a priori, hiperparámetros, número de iteraciones y de cadenas de Markov Monte Carlo, tamaños de muestra efectivo).

### (B) Construcción de la matriz de caracteres funcionales

Un carácter funcional se define como cualquier rasgo fenotípico de un individuo que impacta sobre su éxito reproductivo de forma indirecta a través de sus efectos sobre el crecimiento, supervivencia y reproducción [4]. Como primera medida, se debe reportar la fuente de los caracteres funcionales de las especies estudiadas. Estos pueden provenir de bases de datos, de la literatura, del mismo muestreo realizado en el estudio o de observaciones personales. En el caso de diseños experimentales, estos pueden ser manipulados *a priori* (e.g., altura de los individuos). En todos los casos, debe citarse la fuente de los caracteres. En el caso de las bases de datos en línea, debe citarse la versión y la fecha de descarga, ya que cambian a lo largo del tiempo. La lista de los caracteres utilizados, así como también el tipo de variable (continua, discreta, categórica) y los tamaños de muestra por taxón, deben ser reportados. Debido a que los enfoques de DF asumen una relación entre los caracteres funcionales y su función ecosistémica, esta última debe hacerse explícita para cada carácter, para un claro entendimiento de los procesos ecológicos que se están estudiando. Además, debe indicarse si el carácter es un “carácter respuesta” o “carácter efecto”. Los caracteres respuesta reflejan la respuesta de un organismo al cambio ambiental (e.g., tamaño corporal, tamaño de nidada); una mayor diversidad de caracteres respuesta se espera que ofrezcan mayor resiliencia al sistema contra perturbaciones ambientales. Los caracteres efecto determinan el efecto que un organismo tiene sobre una función ecosistémica (e.g., dieta, comportamiento de forrajeo); una mayor diversidad de caracteres efecto brindarían un mayor rango de funciones ecosistémicas (e.g., polinización, dispersión de semillas). Por último, debe indicarse si el estudio tiene en cuenta el promedio de los individuos para cada taxón, o si se considera la variación intraespecífica de cada uno [5]. En este sentido, la resolución espacial (local, regional, global) es clave para determinar el potencial efecto que pueda tener el promedio y la variación intraespecífica de los caracteres sobre la DF. De forma análoga, la resolución temporal permite establecer si el tiempo puede ser un efecto a considerar en los cambios de la DF (e.g., series temporales de largo plazo, efectos postdisturbio [6]). Por último, debe indicarse si se aplicó alguna transformación a los caracteres (logaritmo, estandarización). Con la información resultante, se construye una matriz  $\mathbf{T}_{s \times p}$  de  $s$  especies  $\times$   $p$  caracteres funcionales.

### (C) Estimación de la DF

Una vez construidas las matrices  $\mathbf{C}$  y  $\mathbf{T}$ , existen diversos enfoques para estimar la DF. Entre los más utilizados se encuentran los índices de DF, que se realiza usualmente en tres pasos: (1) cálculo de una matriz de disimilitud  $\mathbf{D}_{p \times p}$  a partir de la matriz  $\mathbf{T}$ , (2) aplicación de un análisis de agrupamientos o reducción de dimensiones de la matriz  $\mathbf{D}$  [7], y (3) cálculo de índices de DF. Por lo tanto, es indispensable reportar los índices, la medida de disimilitud y la técnica multivariada utilizados [8]. Si se utiliza un método de ordenación, debe indicarse si se realizó alguna corrección de la matriz  $\mathbf{D}$  (en el caso de obtener autovalores negativos), así como el

número de componentes retenidos y el porcentaje acumulado de variación explicada para calcular los índices. A pesar de la gran diversidad de índices presentes en la literatura, se considera que existen cuatro grandes dimensiones de DF (riqueza, equidad, divergencia y redundancia funcional), y que estos índices covarían con alguna de estas dimensiones [9]. Así, debe reportarse la dimensión con la cual se vincula el índice utilizado para una correcta interpretación de los resultados. Asimismo, debe indicarse si estos se calcularon sobre el conjunto de  $p$  caracteres, de cada carácter individual o de subconjuntos de caracteres (que representarían indicadores de diferentes funciones ecosistémicas). Por último, debe indicarse si los índices fueron estandarizados y de qué forma.

#### (D) Ajuste de modelos y reporte de resultados

Para analizar la relación entre DF y diferentes predictores existen dos grandes enfoques: (a) comparación de medidas de DF con distribuciones nulas y (b) relación entre DF y diferentes predictores ambientales (e.g., gradientes de urbanización, uso de suelo, vegetación, competencia). En el caso de los modelos nulos, debe indicarse el tipo de medida del efecto y el número de muestras aleatorias utilizadas. Para el análisis de la FD vs. variables ambientales, se pueden aplicar una gran diversidad de modelos (modelos generalizados, modelos mixtos, modelos aditivos, modelos de ecuaciones estructurales). En estos casos, deben reportarse la distribución del error y función de enlace utilizada, así como el cumplimiento o no de supuestos (heterogeneidad de varianzas, sobredispersión, autocorrelación espacial y temporal). Como medida del efecto, deben reportarse los coeficientes e intervalos de confianza de cada predictor incluido en el modelo, así como la varianza explicada por los predictores ( $R^2$  y derivados). En cada uno de los pasos presentados, deben indicarse el software y versión utilizada, y de ser posible, las bases de datos primarias utilizadas para los análisis. En el mejor de los casos, es siempre recomendable proveer el código utilizado.

## DISCUSIÓN Y CONCLUSIONES

Se espera que este workflow y serie de recomendaciones sean considerados en el futuro por investigadores en el estudio de comunidades y caracteres funcionales. Esta checklist brinda una herramienta para entender y replicar estudios previos, y promueve un punto de partida para adoptar estándares de reproducibilidad en el estudio de la ecología funcional.

## REFERENCIAS

- [1] Baker, M. (2016). "Reproducibility crisis". *Nature*, 533(26): 353-366.
- [2] Cassey, P., y Blackburn, T. M. (2006). "Reproducibility and repeatability in ecology". *BioScience*, 56(12): 958-959.
- [3] Palacio, F. X., Maragliano, R. E., y Montalti, D. (2020). "The costs of ignoring species detectability on functional diversity estimation". *The Auk*, doi: 10.1093/auk/ukaa057
- [4] Violle, C., Navas, M. L., Vile, D., Kazakou, E., Fortunel, C., Hummel, I., y Garnier, E. (2007). "Let the concept of trait be functional!" *Oikos*, 116: 882-892.
- [5] Palacio, F. X., Fernández, G. J., y Ordano, M. (2019). "Does accounting for within-individual trait variation matter for measuring functional diversity?" *Ecological Indicators*, 102: 43-50.
- [6] Palacio, F. X., Ibañez, L. M., Maragliano, R. E., & Montalti, D. (2018). "Urbanization as a driver of taxonomic, functional, and phylogenetic diversity losses in bird communities". *Canadian Journal of Zoology*, 96(10): 1114-1121.
- [7] Laliberté, E., y Legendre, P. (2010). "A distance-based framework for measuring functional diversity from multiple traits". *Ecology*, 91(1): 299-305.



- [8] Palacio, F. X., Apodaca, M. J., y Crisci, J. V. (2020). *Análisis multivariado para datos biológicos. Teoría y su aplicación utilizando el lenguaje R*. 1era Edición, Argentina.
- [9] Mouchet, M. A., Villéger, S., Mason, N. W., y Mouillot, D. (2010). "Functional diversity measures: an overview of their redundancy and their ability to discriminate community assembly rules". *Functional Ecology*, 24(4): 867-876.

## EVALUACIÓN DE ALTERNATIVAS AL USO DE LOS ANTIBIÓTICOS EN ENSAYOS AVÍCOLAS UTILIZANDO MODELOS LINEALES MIXTOS

PALACIOS LUCIANO<sup>1</sup>; FILIPPINI OLGA SUSANA<sup>1</sup>; DE FRANCESCHI MAURICIO<sup>2</sup>

<sup>1</sup>División Estadística, Departamento Cs. Básicas, Universidad Nacional de Luján

<sup>2</sup>Avicultura, Departamento Tecnología, Universidad Nacional de Luján

lucianofepalacios@hotmail.com

### RESUMEN

El aumento de la resistencia a los antimicrobianos, junto con la creciente presión global para reducir o eliminar los antibióticos de la producción de alimentos, ha impulsado un reemplazo por promotores de crecimiento de origen natural, como una alternativa y desafío en la producción avícola. Para confeccionar dietas naturales y libre de sustancias indeseables, el uso de probióticos obtenidos como productos de la fermentación del microorganismo *Bacillus amyloliquefaciens*, favorece el crecimiento de la microbiota beneficiosa e inhibe la patógena, estimula el desarrollo inmunitario y del tracto digestivo, mejorando el índice de conversión alimenticia en pollos parrilleros. Los nuevos enfoques plantean el concepto de nutribiosis, que tiene como objetivo profundizar interacciones entre nutrición-inmunología intestinal-microbiota. Las tecnologías que contengan enzimas, mejoran la digestibilidad y uniformidad del lote, conjuntamente con probióticos para proteger desafíos de salud y acelerar el sistema inmunológico, optimizando la salud intestinal y el rendimiento productivo. El objetivo de este trabajo es comparar la respuesta productiva de pollos parrilleros frente la presencia y ausencia de antibióticos en sus dietas. Las variables productivas estudiadas, fueron peso semanal, consumo semanal y acumulado de alimento, conversión alimenticia y ganancia diaria. El ensayo fue realizado siguiendo un Diseño en Bloques Completos al azar (DBCA) considerando 4 tratamientos y 20 repeticiones. A: Dieta tradicional con antibiótico, B: Dieta tradicional sin antibiótico, C: Dieta con subproducto de la fermentación sin antibiótico y D: sin antibiótico con dieta a base de prebiótico y enzimas. Se consideró para el procesamiento de los datos, un Modelo Lineal Mixto considerando los tratamientos y el tiempo como efectos fijos y su interacción, los bloques y corrales como efectos aleatorios. En general, se observó que el tratamiento con probióticos es el de mejor performance en peso y ganancia de peso, está entre los de menor consumo y es el más eficiente en conversión.

**Palabras Clave:** *pollos parrilleros, probióticos, dietas naturales, inmunidad, promotores del crecimiento, nutribiosis, microbiota, digestibilidad.*

### INTRODUCCIÓN

Existe un movimiento global para reducir el uso de antibióticos, tanto terapéuticos como promotores del crecimiento (ACP), para superar la resistencia a los antimicrobianos y preservar la eficacia de los antibióticos disponibles (Abudabos *et al.*, 2015; Araujo *et al.*, 2019). Los patógenos están aumentando su incidencia debido a la reducción o eliminación de antibióticos de las dietas en la producción avícola, como consecuencia de la presión de los consumidores de carne y nuevos paradigmas en la alimentación. La exposición diaria a patógenos como *Escherichia coli*, *Clostridium perfringens* y *Salmonella* puede provocar deficiencias de rendimiento en los lotes de producción avícola extensivos. La energía que las aves podrían

dirigir hacia el crecimiento ahora debe usarse para mantener la salud y superar estos desafíos patógenos. Muchos probióticos que se emplean en dietas para aves comerciales, tanto en pollos parrilleros, como gallinas de postura y otras especies, tienen el objetivo de favorecer el crecimiento de la microbiota benéfica, frenando el desarrollo de la más patógena, estimulando el desarrollo inmunitario del animal y principalmente, estimulando el desarrollo del tracto digestivo (Da Silva *et al.*, 2019). Como resultado de sus efectos combinados, mejora la salud gastrointestinal e inmunitaria del animal, resultando en una mejora del índice de conversión alimenticia en animales sanos y en aquellos infectados, incluso en animales asintomáticos. Muchos de los probióticos surgen de la fermentación de microorganismos, tal es el caso de *Bacillus amyloliquefaciens* (Abudabos *et al.*, 2015; Araujo *et al.*, 2019).

En la actualidad, diversos estudios en pollos parrilleros han puesto a prueba la combinación de diferentes probióticos y enzimas simultáneamente, con beneficios sinérgicos en dietas sin antibióticos. La combinación de enzima/prebiótico ofrece un retorno económico mayor, y mejora en los parámetros zootécnicos, debido a una mejor digestibilidad y salud intestinal. Además, se mejoran de rendimiento al utilizar fitasas, xilanasas, amilasas, proteasas con probióticos de *Bacillus*, en lugar de un promotor de crecimiento antibiótico (AGP). El objetivo de este trabajo es comparar la respuesta productiva de pollos parrilleros frente la presencia y ausencia de antibióticos en sus dietas.

## MATERIALES Y MÉTODOS

El ensayo fue conducido bajo un Diseño en Bloques Completos Aleatorizados, considerando 20 bloques. Se evaluaron cuatro tratamientos dentro de cada bloque. Los bloques permitieron controlar diferencias de temperatura ambiental, presión del agua en bebederos y condiciones de manejo. El galpón está dividido en 80 corrales, cada uno de 1 m<sup>2</sup>. Se alojaron 12 aves machos por corral. Se consideró que los tanques de agua no introducen variabilidad y cada uno abastecerá la línea más cercana de corrales. Los tratamientos fueron: A-Dieta tradicional con antibiótico promotor del crecimiento APC (100 g/ton); B- Dieta tradicional sin APC, C- B + probióticos de tres cepas *Bacillus amyloliquefaciens*; D- B+ prebiótico de tres cepas *Bacillus amyloliquefaciens* y enzimas xilanasas, amilasa y proteasa. Las variables que se relevaron en intervalos de tiempo de 7 días fueron, peso promedio semanal, consumo semanal y acumulado de alimento, conversión alimenticia y ganancia diaria de peso durante 6 semanas.

Los datos por semanas se procesaron con un Modelo Lineal, para un diseño en Bloques Completos al Azar y las Comparaciones Múltiples se realizaron según los Métodos Clásicos, Pruebas de Bonferroni y Tukey. También se realizaron pruebas robustas (Welch) para las comparaciones entre grupos cuando fallaron los supuestos de homogeneidad de varianzas, con un Análisis de Variancias para un diseño en Bloques.

Se consideró para el procesamiento de los datos, un Modelo Lineal Mixto, considerando los tratamientos, el tiempo y su interacción como efectos fijos, los bloques y corrales (efecto de parcela) como efectos aleatorios.

## RESULTADOS Y DISCUSIÓN

Los resultados indican que el tratamiento C con la inclusión de probióticos en la dieta, es el de mejor performance en peso y ganancia de peso, está entre los de menor consumo y es el más eficiente en cuanto conversión.

La dieta con probióticos (C) presenta mejor peso en las aves que el resto de los tratamientos, mientras que a los 42 días se comporta similarmente a la opción A.

Los individuos que recibieron los tratamientos A y C manifestaron el mayor consumo en las etapas finales del ensayo. Respecto a la conversión alimenticia, a partir de los 35 días de edad el tratamiento C tiene el mejor índice hasta concluir el ensayo. En la última medición del ensayo, los tratamientos A, B y C demostraron las mejores ganancias respecto a la dieta D (Tabla 1).

**Tabla N°1: Variables zootécnicas medidas**

Tratamiento	Tiempo	Peso	Ganancia de peso diaria	Consumos acumulados	Conversión alimenticia
	[días]	[gs]	[gs]	[gs]	
A	14	370,821 b	31,518 a	463,825 a	1,305 a
	21	735,893 b	52,153 b	1115,875 a	0,886 a
	28	1202,941 b	65,936 b	2128,420 a	1,772 a
	35	1608,739 b	57,971 ab	3101,485 b	1,932 b
	42	2192,787 bc	83,435 b	4330,175 d	1,977 b
B	14	348,179 a	30,736 a	459,921 a	1,344 a
	21	679,323 a	47,277 a	1090,703 a	0,927 a
	28	1109,664 a	61,885 b	2082,639 a	1,880 b
	35	1559,104 b	64,606 ab	3058,138 ab	1,965 b
	42	2114,70 b	79,372 b	4227,711 bc	2,002 b
C	14	368,429 b	30,231 a	456,546 a	1,351 ab
	21	729,286 b	51,551 b	1100,425 a	0,883a
	28	1194,524 b	66,451 b	1100,425 a	1,768a
	35	1666,931 c	67,487 b	3101,941 b	1,864a
	42	2203,505 c	76,653 b	4280,506 cd	1,946a
D	14	348,107 a	28,612 a	461,191 a	1,415 b
	21	687,143 a	48,4344 b	1108,255 a	0,941 a
	28	1084,034 a	56,963 a	2055,622 a	1,899 b
	35	1417,745 a	47,973 a	2945,935 a	2,087 c
	42	1903,745 a	69,336 a	4057,856 a	2,139 c

Letras distintas por columna en cada tiempo indican diferencias estadísticamente significativas (valor  $p < 0.05$ )

## REFERENCIAS

- [1] Di Rienzo, J. A.; Casanoves, F. y Macchiavelli, R. E. (2011) “*Modelos lineales mixtos: aplicaciones en InfoStat*”. 1a. ed. - Córdoba: Grupo Infostat, 2011. 193 p.: il. ISBN 978-987-27045-0-6. Argentina.
- [2] Di Rienzo, J. A.; Macchiavelli, R. E. y Casanoves, F. (2017) “*Modelos lineales generalizados mixtos: aplicaciones en InfoStat*.” 1a. ed. - Córdoba: Grupo Infostat, 2017.il. ISBN 978-987-42-4985-2. Argentina.
- [3] Da Silva Raposo, R.; Defensor, R. y Grahl, T. (2019). “*Uso de probióticos na avicultura para o controle da Salmonella spp.: Revisão e perspectivas de utilização*”. Pubvet, vol 13 (N° 4, a305): pp1-8.
- [4] Abudabos, A. M.; Al-Batshan, H. A. y Murshed, M. A. (2015). “Effects of prebiotics and probiotics on the performance and bacterial colonization of broiler chickens”. South African Journal of Animal Science, vol 45(N°4):pp419-428.
- [5] Araujo, R.; Barbieri, A.; Silva, KM.; G Ventura, G. y Polycarpo, VC. (2019). “*Performance and economic viability of broiler chickens fed with probiotic and organic acids in an attempt to replace growth promoting antibiotics*”.Braziliam Journal Poultry Science. vol.21 no.2 Campinas 2019. Epub July 29, 2019. Versión obtenida el 01/09/2020. <https://doi.org/10.1590/1806-9061-2018-0912>

## ANÁLISIS ECONÓMETRICO DE LA VALORACIÓN DE ACEITE DE OLIVA VIRGEN EXTRA<sup>8</sup>

ROLDÁN CAMILA<sup>a</sup>, LUPÍN BEATRIZ<sup>a</sup>, TEDESCO LORENA<sup>b</sup>

<sup>a</sup>Grupo de Investigación Economía Agraria, FCEyS-UNMdP

<sup>b</sup>IIESS-CONICET, Departamento de Economía-UNS

camila.anto.rolدان@gmail.com, beatrizlupin@gmail.com, ltedesco@iiess-conicet.gob.ar

### RESUMEN

En el presente Trabajo, se estudian las preferencias y las elecciones de los consumidores de aceite de oliva virgen extra residentes en la Ciudad de Bahía Blanca a fin de estimar la valoración de determinados atributos de calidad, centrandose el interés en el producido en el Sudoeste de la Provincia de Buenos Aires. Se emplearon datos provenientes de una encuesta, con representatividad censal, que incluyó un *Choice Modelling*. Dicho relevamiento, comprendió 223 casos y fue llevado a cabo en la Ciudad mencionada, durante los meses de noviembre y diciembre del año 2017. Se aplica un *Alternative Specific Logit Model*, de efectos principales, que permite conocer la contribución de ciertos atributos –origen geográfico, sabor, material del envase y precio– a la utilidad de los consumidores y su disposición a pagar por los mismos. Los resultados generales indican que los consumidores valoran un aceite de oliva virgen extra proveniente del Sudoeste Bonaerense, de sabor suave y con envase de vidrio, siendo el segundo atributo el que mayor utilidad les reporta y, en consecuencia, se encuentran dispuestos a pagar un diferencial de precio mayor por él.

**Palabras clave:** *alternative specific logit model, choice modelling, consumidores, calidad percibida, atributos de calidad, sudoeste bonaerense.*

### INTRODUCCIÓN

Durante los últimos años, en los mercados alimentarios, ha cobrado relevancia el papel que tiene el consumidor debido a su preocupación por la calidad, creciente interés por adoptar una dieta y hábitos saludables, concientización ambiental e inclinación por los productos provenientes de las zonas cercanas a donde habita, entre otras cuestiones. Por su parte, la industria enfrenta el desafío de innovar cada vez más dado el comportamiento de los consumidores pero, también, a la mayor competencia promovida por la globalización.

En este contexto, el aceite de oliva virgen extra (AOVE) se ha posicionado como un alimento de calidad diferenciada, conocido por sus propiedades para prevenir determinadas enfermedades. Respecto a nuestro país, una de las regiones con mayor desarrollo en la materia es el Sudoeste Bonaerense (SOB), donde se encuentra un núcleo de productores que han sabido aprovechar distintas ventajas agroecológicas y de servicios para elaborar un AOVE de elevada calidad objetiva. De esta manera, a los requerimientos del consumidor, se suma el interés de los productores del SOB por promocionar sus aceites a través de la identidad territorial. Es por ello que iniciaron la evaluación de la construcción de un sello de calidad territorial, siendo esencial conocer la valoración de los consumidores por el origen geográfico del producto y otros

---

<sup>8</sup>Este trabajo se basa en la tesis para acceder al grado de Licenciada en Economía (FCEyS-UNMdP) de Camila Roldán, bajo la dirección de la Lic. Beatriz Lupín (UNMdP) y la codirección de la Lic. (Dra.) Lorena Tedesco (UNS), defendida en marzo 2020.

atributos propios. Siguiendo esta línea, el objetivo del Trabajo es analizar la valoración de determinados atributos del AOVE por parte de los consumidores residentes en la Ciudad de Bahía Blanca, con énfasis en la disposición a pagar (DAP) por aquellos que lo distinguen del de otras procedencias geográficas. Las hipótesis planteadas son: H1) *Para los consumidores, el “origen geográfico” es el atributo de mayor importancia relativa* y H2) *Los consumidores están dispuestos a pagar un adicional por AOVE regional.*

## MATERIALES Y MÉTODOS

Para la realización de este estudio, se contó con datos provenientes de una encuesta sobre consumo de aceite de oliva, que incluyó un Experimento de Elección –*Choice Modelling*–. Participaron 223 consumidores del producto, de 18 años y más, con decisión en la elaboración y/o compra de alimentos de sus hogares. Dichos participantes, fueron captados en comercios de la Ciudad de Bahía Blanca<sup>9</sup>, durante los meses de noviembre y diciembre del año 2017, conforme cuotas censales de sexo y edad y cubriendo distintas realidades socioeconómicas. Los atributos evaluados fueron: ORIGEN GEOGRÁFICO (ORIG) –2 niveles: SOB, otra procedencia (del país o del exterior)–, SABOR (SAB) –2 niveles: suave y fuerte/intenso–, ENVASE (ENV) –2 niveles: plástico y vidrio– y PRECIO (PRE) –3 niveles: \$ 150, \$ 175 y \$ 200, el litro<sup>10</sup>. Estos atributos fueron dispuestos de manera de conformar 9 alternativas del producto, siguiendo un diseño fraccional ortogonal –Algoritmo *Orthoplan* (Software IBM®SPSS®)–, a los que se agregó la alternativa “no compraría ninguna de las alternativas expuestas” (*opt-out*)<sup>11</sup>.

A fin de estimar los efectos principales y la DAP, se aplicó un *Alternative Specific Logit Model* (Paquete ASCLOGIT del Software Stata®). Se trata de un caso específico del *Conditional Logit Model* (McFadden, 1973) que permite combinar características de cada alternativa y de cada participante (Arela Bobadilla, 2016). Cabe mencionar dos cuestiones metodológicas relevantes –la codificación de las variables y la inclusión de Constantes por Alternativas Específicas (ASC)– y una operativa –armado de la base de datos para la estimación econométrica–. En este Trabajo, se optó por codificar todas las variables como *dummy*, a excepción del PRE que fue tratada como una variable cuantitativa discreta; se incorporó una ASC para la *opt-out* y se adoptó el formato *long form*<sup>12</sup>.

## RESULTADOS Y DISCUSIÓN

El modelo expuesto seguidamente es adecuado tanto desde la perspectiva econométrica como económica. Por ende, los resultados derivados pueden servir de orientación para los productores y otros agentes del Sector y para los encargados de diseñar políticas públicas referidas a alimentación saludable y a la promoción de las economías regionales.

<sup>9</sup>Se realizó la experiencia en esta Ciudad pues la misma se distingue por su superficie, población e importancia relativa como proveedora de bienes y servicios y nodo de comunicaciones y transporte a nivel regional y nacional. Asimismo, constituye un punto de interés para los productores regionales a fin de promover y comercializar sus productos.

<sup>10</sup>Los atributos fueron seleccionados conforme una encuesta exploratoria realizada en el año 2015, entrevistas a informantes calificados –productores, comercializadores e investigadores agrónomos, biólogos y químicos–, grupos focales y una exhaustiva revisión bibliográfica.

<sup>11</sup>La experiencia se sustanció en el Proyecto “Desarrollo del Olivo del SOB” –Programa Universidades Agregando Valor, SPU-Ministerio de Educación y Deportes, convocatoria 2016, bajo la dirección de la Lic. (Dra.) Lorena Tedesco (UNS) y coordinación de la Lic. Beatriz Lupín (UNMdP)–. Para ampliar al respecto, se sugiere la lectura de Roldán (2020).

<sup>12</sup>Para ampliar al respecto, se sugiere la lectura de Roldán (2020).

Estos resultados indican que las variables explicativas, que representan a los atributos considerados, presentan los signos esperados según la Teoría Económica, además son estadísticamente distintas a 0 –se rechaza la  $H_0$ ) de la Prueba de Wald que establece *La probabilidad de éxito es independiente de la variable explicativa*–. Por su parte, la bondad del ajuste del Modelo es satisfactoria. (Tabla 1).

La procedencia del “SOB” (ORIG), el sabor “suave” (SAB) y el envase de “vidrio” (ENV) favorecen la elección de un AOVE con cada uno de estas características, siendo la segunda la que más contribuye a la utilidad del participante –no se verifica la  $H_1$ )–. Asimismo, se cumple la Ley de la Demanda pues a medida que el precio (PRE) disminuye, aumenta la utilidad.

Tabla 1: Estimación del ASLM de Efectos Principales

–Muestra total: 223 casos, 2.676 observaciones–

Variables explicativas	Tipo de variable/Escala		Coefficientes estimados	Errores estándar	Estadístico Z de Wald	
<b>ORIG</b>	1: Originario del SOB	CUALI	0,4756334***	0,1020596	5,35	
<b>SAB</b>	1: Suave	CUALI	0,7724321***	0,1019778	7,57	
<b>ENV</b>	1: Vidrio	CUALI	0,4382286***	0,0888852	4,29	
<b>PRE</b>	Discreta: \$ 150, \$ 175, \$ 200		CUAN TI	-0,0128995***	0,0029176	-4,42
<b>ASC<sub>OUT</sub></b>	1: Elige la alternativa <i>opt-out</i>		CUALI	-3,5248290***	0,5292180	-6,66

Wald ( $\chi^2$ ) = 205,34; \*\*\* significación estadística al 1%. Probabilidad  $> \chi^2 = 0,000 < 1\%$ ; Log Likelihood = -757,26752.

Por último, la  $ASC_{OUT}$  es estadísticamente significativa y su signo negativo implica que los participantes tienen mayor probabilidad de comprar una de las 9 alternativas de producto que de no comprar ninguna de ellas. Dado que el atributo monetario (PRE) es estadísticamente significativo y tiene el signo correcto, fue posible estimar la DAP por cada uno de los restantes atributos (ORIG, SAB, ENV):

Tabla 2: DAP por atributos del AOVE

–Muestra total: 223 casos, 2.676 observaciones–

Atributo –nivel–	Media DAP –\$/l–	Errores estándar	Intervalos de confianza
<b>ORIG</b> SOB	36,87***	11,27428	[14,77 - 58,96]
<b>SAB</b> Suave	59,88***	13,88381	[32,66 - 87,09]
<b>ENV</b> Vidrio	33,97***	7,625181	[19,02 - 48,91]

\*\*\* significación estadística al 1%.

De la tabla anterior, se desprende que, en promedio, los participantes están dispuestos a pagar un adicional de \$ 36,87, \$ 59,88 y \$ 33,97 por litro de AOVE procedente del SOB –se verifica la  $H_2$ )–, con sabor suave y en envase de vidrio en comparación de lo que pagarían por uno con los otros niveles considerados, *ceteris paribus*. Estos valores, representan el cambio necesario en el precio del producto para compensar el cambio en el origen geográfico, en el sabor y en el material del envase, es decir, la tasa marginal de sustitución entre cada uno de dichos atributos y el precio.

## REFERENCIAS

[1] Arela Bobadilla, R. W. (2016). *Análisis microeconómico del impacto de la informalidad laboral en el déficit cuantitativo tradicional de vivienda y la decisión entre alquiler y comprar una vivienda en el Perú entre 2012 y 2015*. (Tesis de Grado). Facultad de Ciencias Económico Empresariales-Universidad Católica San Pablo, Arequipa-Perú.



- [2] Hensher, D. A.; Rose, J. M. & Greene, W. H. (2007). *Applied Choice Analysis. A primer*. New York-USA: Cambridge University Press.
- [3] Louviere, J. J.; Hensher, D. A. & Swait, J. D. (2000). *Stated Choice Methods. Analysis and applications*. New York-USA: Cambridge University Press.
- [4] McFadden, D. (1973). Conditional Logit Analysis of Qualitative Choice Behavior. Publicado en Zarembka, P. (Ed.), *Frontiers in Econometrics*, Academic Press, 105-142.
- [5] Roldán, C. (marzo 2020). *Disposición a pagar por atributos de calidad de aceite de oliva virgen extra en la Ciudad de Bahía Blanca-Argentina*. (Tesis de Grado). FCEyS-UNMdP, Mar del Plata-Argentina.
- [6] Train, K. E. (2009). *Discrete Choice Methods with simulation*. New York-USA:Cambridge University Press.

## MODELO BINOMIAL NEGATIVA PARA AJUSTAR LA PRESENCIA DE CHIMANGOS EN EL AEROPUERTO INTERNACIONAL DE LA CIUDAD DE NEUQUÉN

RUBIO NATALIA<sup>1</sup>, ANGELICCHIO JUAN<sup>2</sup>, GIMENEZ GUSTAVO N.<sup>1</sup>

<sup>1</sup>Departamento de Estadística, Facultad de Economía y Administración UNCo

<sup>2</sup>Facultad de Ciencias de Ciencias del Ambiente y la Salud, Unco  
rubionatalia@yahoo.com.ar

### RESUMEN

Una problemática que afecta al tráfico aéreo de los aeropuertos nacionales e internacionales es la presencia de fauna silvestre principalmente de aves. Los seres humanos empezaron a compartir con ellas el espacio aéreo hace apenas aproximadamente 100 años. Desafortunadamente, cuando aeronaves y aves intentan usar al mismo tiempo un espacio aéreo, se dan colisiones entre ellos (Cleary & Dolbeer, 2005). El impacto entre aeronaves y fauna silvestre se ha convertido en una preocupación importante a nivel internacional debido a que genera serios problemas económicos y de seguridad pública (Martínez Moreno, 2019). En el Aeropuerto de Neuquén una de las especies de aves de mayor riesgo es el chimango (Milvago chimango). Para mitigar este riesgo resulta fundamental conocer por qué la frecuencia de avistajes de aves se incrementa, lo que implica analizar patrones de conducta frente a características ambientales. Para esta investigación se registró la frecuencia de chimangos semanalmente mediante el método de conteo por puntos durante 1 hora al amanecer y las principales variables climáticas como velocidad del viento, nubosidad, época del año, temperatura máxima y mínima del día, humedad y precipitación promedio. Por lo tanto considerando la variable de respuesta cantidad de Chimangos avistados por día, se hizo uso de los Modelos Lineales Generalizados, aplicando en una primera instancia Un modelo Poisson y luego un Modelo Binomial Negativo. Comenzando el análisis con un modelo saturado y se realizó una selección de variables mediante el algoritmo Step con criterio BIC. Luego se comparó tanto la Deviance Residual como el valor del AIC de los modelos seleccionado con ambas distribuciones. Resultando que el modelo Binomial Negativo es el que mejor explica la presencia de chimangos en función la estación del año, Velocidad del Viento y Estación del Año.

**Palabras clave:** *modelo lineal generalizado, modelo binomial negativo, modelo poisson.*

### INTRODUCCIÓN

La naturaleza y magnitud de los problemas que enfrenta un aeropuerto dependen no solo del volumen de tráfico aéreo, sino también de la presencia de fauna silvestre. La presencia de cualquier ave constituye un peligro que se agrava tanto por el tamaño del ave como por el número de individuos que vuelan juntos. En el Aeropuerto de Neuquén una de las especies de aves de mayor riesgo es el chimango (Milvago chimango). Para mitigar este riesgo resulta fundamental conocer por qué la frecuencia de avistajes de aves se incrementa, lo que implica analizar patrones de conducta frente a características ambientales. Se registró la frecuencia de chimangos semanalmente mediante el método de conteo por puntos durante 1 hora al amanecer y las principales variables climáticas como velocidad del viento, nubosidad, época del

año, temperatura máxima y mínima del día, humedad y precipitación promedio. Por lo tanto considerando la variable de respuesta cantidad de Chimangos avistados por día, se aplicó un modelo Poisson y un Modelo Binomial Negativo. Mediante el proceso de selección de variables, el análisis de los residuos y comparando los modelos mediante el uso de Test de verosimilitud se arribó a que la mejor distribución para las observaciones es el de distribución Binomial negativa, que explica la presencia de los chimangos en función de la Velocidad del Viento y Estación del Año.

## **MATERIALES Y MÉTODOS**

El Aeropuerto Internacional Presidente Juan Domingo Perón se encuentra ubicado en la ciudad de Neuquén, provincia del Neuquén, aproximadamente a 6 kilómetros del centro de la ciudad en el sector oeste entre las coordenadas 38° 57' 11" de latitud norte y 68° 08' 00" longitud oeste. Ocupa un predio de aproximadamente 254 ha casi de manera rectangular, es considerado uno de los aeropuertos más importantes dentro de la Patagonia argentina. Las áreas internas del aeropuerto presentan diferentes tipos de vegetación, desde sitios con altas densidades de árboles y matorrales en el sector periférico del predio, hasta pastizales en el sector que rodea la pista.

El aeropuerto está rodeado por áreas urbanas y suburbanas con diferentes usos del suelo, como por ejemplo: loteos, barrios, asentamientos, tierras dedicadas a la agricultura y a la ganadería, lagunas artificiales, canales de riego y ríos, como lo es el río Limay.

Metodología de trabajo: En la primera instancia se realizó avistaje de aves en el Aeropuerto Internacional JD Perón. Para esto se llevó a cabo el método de conteo por puntos, uno de los métodos más comunes para obtener información sobre composición, abundancia relativa y densidad de las especies detectadas de forma visual y auditiva, en el cual el observador establecerá un punto fijo en la pista de aviones donde se para y cuenta el número de chimangos Milvago que se observen en la pista y/o que pasen sobrevolando la misma.

El relevamiento se realizó una vez por semana, durante un año completo. Así mismo se tomó nota de las condiciones climáticas del momento del avistaje (temperatura, velocidad del viento, dirección del viento, nubosidad, etc.); registrando también el número de despegues y aterrizajes de aviones que ocurran durante el relevamiento de cada día.

Análisis de datos: Se realizó una modelización mediante el uso de los Modelos Lineales Generalizados. La modelación de datos experimentales desde el marco teórico de los modelos lineales generalizados brinda la posibilidad de analizar datos con estructuras de dependencia, desbalances y falta de normalidad o simplemente datos donde la variable de respuesta tiene una distribución Poisson, Binomial o Gamma, entre otras. Estos permiten contemplar la falta de cumplimiento de los supuestos tradicionales y modelar, de manera flexible, complicadas estructuras de datos. Existen muchos beneficios que pueden ser obtenidos con el uso de modelos generalizados. En algunas situaciones se incrementa la precisión de las estimaciones; en otras, se amplía el espacio de inferencia y se comprende mejor la estructura de los datos.

Debido a que la variable de respuesta es de naturaleza Poisson, se aplicó un modelo lineal generalizado, partiendo de un modelo saturado que incluye todas las variables explicativas y sus interacciones para luego arribar a un modelo más parsimonioso.

Una vez obtenido el modelo que mejor ajustó los datos se realizó el mismo procedimiento pero con distribución Binomial Negativa. Finalmente se compararon ambos modelos, para luego realizar las interpretaciones de los coeficientes estimados en el contexto del problema. Para llevar a cabo todo el análisis se utilizó el software libre R.

## RESULTADOS Y DISCUSIÓN

El primer modelo ajustado fue un modelo saturado con todas las interacciones considerando un Modelo Lineal Generalizado con distribución Poisson al cual se le realizó un proceso de selección de variables obteniendo un modelo que tiene en cuenta las interacciones dobles entre la Humedad la Velocidad del Viento y la Estación del año. Se realizó el mismo procedimiento pero con Distribución Binomial Negativa obteniendo los resultados que se muestran en la tabla 1.

Pudiendo observar que no cabe duda que el mejor modelo considere una distribución Binomial Negativa.

Ahora considerando dicha distribución al comparar los modelos con interacción versus el modelo que considera solo los efectos principales se obtuvo los resultados que se muestran en la tabla 2:

Tabla 1: Comparación Modelo Poisson Binomial negativo:

	<b>Modelo Poisson</b>		<b>Modelo Binomial Negativo</b>	
	<b>Efectos Principales:</b> H – VV - Est	<b>Interacciones Dobles</b> entre: H – VV - Est	<b>Efectos Principales:</b> H – VV - Est	<b>Interacciones Dobles</b> entre: H – VV - Est
<b>Deviance Residual</b>	1458.9 con 37gl	1225.7 con 30 gl	44.449 con 37 gl	44.20 con 30 gl
<b>AIC</b>	1761.9	1542.7	496.88	503.38

Tabla 2: Test de Verosimilitud

Model	df	2 x log-lik.	Test	df	Pr(Chi)
H + VV + E	37	-482.8779			
H * VV + H*E + VV * E	30	-475.3784	1 vs 2	7	0.3787916

Por lo tanto podemos destacar dos grandes conclusiones. Primero que la distribución es la más adecuada es la Binomial Negativa, la cual explica de manera más eficiente a la aparición de chimangos en función de las variables climáticas. Y la segunda es que las variables climáticas que mejor explican a la presencia de estas aves es la Humedad, la estación del año y la velocidad del viento.

También podemos afirmar que la cantidad de chimangos avistados por día es mayor tanto en las estaciones de Verano como en la de Otoño respecto de la de Invierno. Y de manera contraria ocurre con la humedad y la velocidad del viento, es decir que a medida que aumentan dichas condiciones la cantidad de Chimangos disminuye (Tabla 3).

Tabla 3. Estimación de los coeficientes.

Coeficientes	Estimación	Error Estándar	Pr(> z )
Intercepto	5.956380	0.093253	< 2e-16 ***
Hum	-0.018643	0.001020	< 2e-16 ***
VelViento	-0.023550	0.002165	< 2e-16 ***



EstaciónOtoño	0.774365	0.044911	< 2e-16 ***
EstaciónPrimavera	0.373287	0.049182	3.2e-14 ***
EstaciónVerano	0.865801	0.043829	< 2e-16 ***

## REFERENCIAS

- [1] Barrientos, C., González-Acuña, D., Moreno, L., & Ardiles, K. (2016). Aves asociadas al Aeropuerto Carriel Sur de Talcahuano, sur de Chile: evaluación de peligro aviario. *Gayana (Concepción)*, 80(1), 40-55.
- [2] De la Peña, M. R., & Salvador, S. (2019). Aves argentinas: descripción, comportamiento, reproducción y distribución. Charadriidae a Trochilidae. *Comunicaciones del Museo Provincial de Ciencias Naturales "Florentino Ameghino" (Nueva Serie)*.
- [3] Zuur A. F., Hilbe J. M., Leno E. N. (2013). *A Beginner's Guide to GLM and GLMM with R: A Frequentist and Bayesian Perspective for Ecologists*. Highland Statistics Ltd.
- [4] Bolker R. B. M. (2007). *Ecological Models and Data* in. Princeton University Press.
- [5] Dobson, A. J. (1990). *An Introduction to Generalized Linear Models*. London: Chapman and Hall.

## **ESTUDIO DEL RENDIMIENTO ACADÉMICO DE ESTUDIANTES DE CONTADOR PÚBLICO NACIONAL**

SABINO GUILLERMO<sup>1</sup>, ARENAS LUIS<sup>1</sup>

<sup>1</sup>*Departamento de Estadística, Facultad de Economía y Administración, UNCo  
agasabino@gmail.com*

### **RESUMEN**

El rendimiento académico de los estudiantes universitarios en sus primeros años de carrera depende de múltiples motivos. Se ve afectado por factores socioeconómicos, familiares, emocionales, como así también por la formación previa recibida, capacidad de adaptación a las nuevas metodologías de enseñanza-aprendizaje, entre otras. El acceso a la información académica necesaria para explicar este fenómeno en el sistema universitario argentino es el sistema SIU Guaraní. Al ser una problemática multi causal, la base de datos está conformada por múltiples "tablas crudas" de datos, a las que se pueda acceder mediante consultas a través del Sistema SIU Wichi. El objetivo principal del trabajo es realizar una aproximación al estudio de la problemática mediante la obtención de una primer base de datos; y en segunda instancia abordar este conjunto de datos mediante diversos métodos estadísticos exploratorios. En esta ocasión, nos concentramos únicamente en los estudiantes de Contador Público Nacional de la Facultad de Economía y Administración de la UNCo (Cohortes 2009 – 2017) que han atravesado con éxito los dos primeros años de la carrera, esto corresponde al 10% de la matrícula total. Se analizó el rendimiento académico en función de variables que indican la dificultad de aprobar las 13 primeras asignaturas del plan de estudios. Se evidenció que el tiempo que los estudiantes toman entre cursar una materia y rendir sus exámenes finales repercute no solo en la acreditación de los espacios curriculares, sino también en la calidad de sus calificaciones. Entendiendo que la problemática es compleja, se prevé a futuro continuar la línea de trabajo incorporando estudiantes que abandonaron sus estudios, como así también otros trayectos académicos y variables socioeconómicas que repercuten en el rendimiento final.

**Palabras clave:** *sobrevida, regresión logística, análisis de correspondencias.*

### **INTRODUCCIÓN**

El rendimiento académico de los estudiantes es un componente clave para determinar si una institución está alcanzando sus objetivos educativos. La existencia de un programa de evaluación para documentarlo es esencial para elaborar propuestas de mejora. Existen diferentes aspectos que se asocian al rendimiento académico, entre los que intervienen componentes tanto internos como externos al individuo. Pueden ser de orden social, cognitivo y emocional, que se clasifican en tres categorías: determinantes personales, determinantes sociales y determinantes institucionales (Garbanzo Vargas, 2007). Al tratarse de una primera aproximación a la problemática, este trabajo aborda una fracción específica de un determinante institucional que afecta directamente sobre la administración académica de los estudiantes: el momento adecuado para rendir su examen final. El estudio exhaustivo de esta componente, permitirá rediscutir la confección del calendario académico estudiantil, adecuaciones en el sistema de promoción, plazo para rendir finales, entre otras acciones.

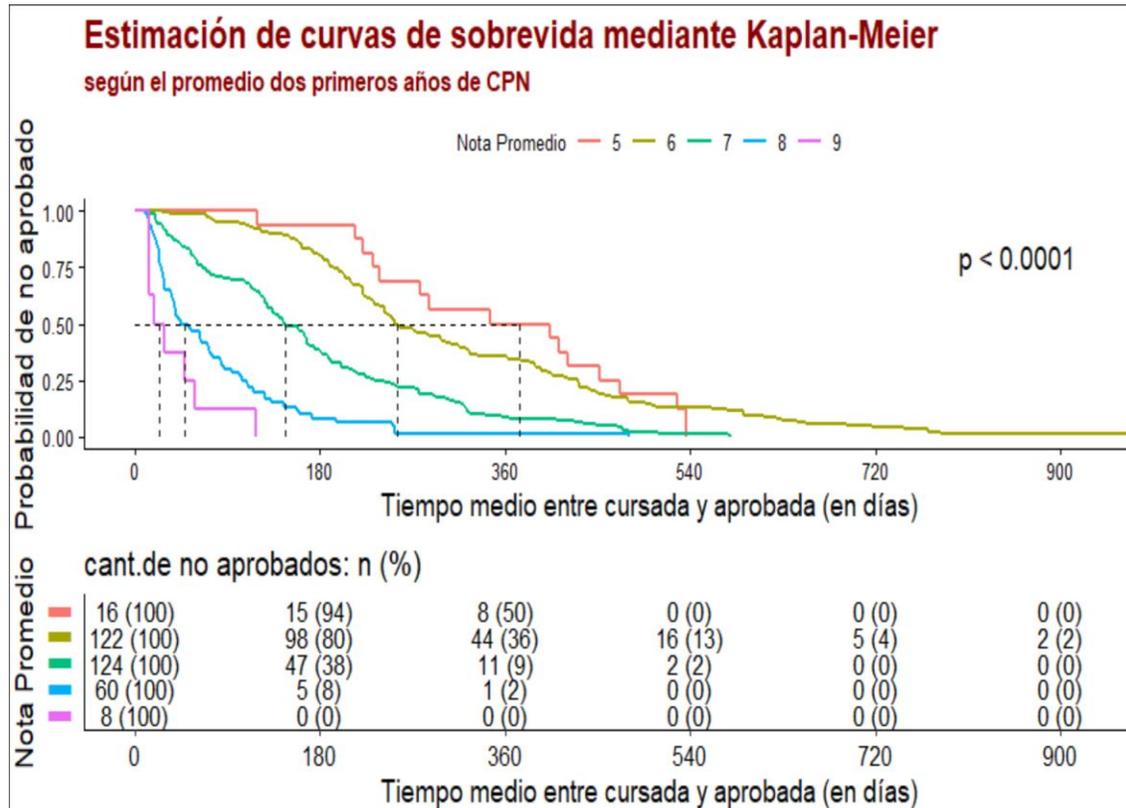
## **MATERIALES Y MÉTODOS**

Para poder acceder a los rendimientos académicos de los estudiantes, se utilizó el sistema SIU-Wichi, que está desarrollado sobre la plataforma Pentaho. Se tomó como población objetivo los 330 estudiantes de Contador Público Nacional de la Facultad de Economía y Administración de la UNCo (Cohortes 2009 – 2017) que aprobaron las primeras trece materias del plan de estudios. Mediante la herramienta Saiku se accedió, para cada materia, al tiempo entre que inició los estudios y cursó, tiempo entre que cursó y aprobó, nota final, número de intentos hasta aprobar, tiempo que tardó en rendir por primera vez y si aprobó o no en esa primer instancia.

La estrategia inicial para abordar el problema, fue mediante métodos multivariados como el Análisis Factorial Múltiple. El gran número de variables asociadas, sumado al número de individuos, requiere de un análisis más extenso, por lo que en esta primera etapa se decidió utilizar metodologías bivariadas. Con el objetivo de analizar el tiempo necesario para aprobar, se realizaron asociaciones con la calificación final (nota) mediante curvas de Kaplan-Meier (Kalbfleisch, 2002) utilizando los paquetes survival (Therneau, 2020) y survminer (Kassambara, 2020) del ambiente R; mientras que para cuantificar la asociación con la asignatura, se utilizó un Análisis de Correspondencias. Por otra parte, para determinar la influencia de la asignatura y el tiempo necesario para preparar el final sobre la posibilidad de acreditar la materia en primera instancia se recurrió a la regresión logística.

## **RESULTADOS Y DISCUSIÓN**

La siguiente figura presenta las curvas de sobrevivencia de Kaplan-Meier las cuales muestran la probabilidad estimada de no haber aprobado en función del tiempo promedio hasta aprobar (las líneas punteadas representan las medianas). Cada curva representa la calificación promedio de las trece materias. Se puede apreciar que a medida que disminuye la nota promedio, aumenta el tiempo medio requerido para aprobar, por ejemplo, la mitad los alumnos con promedio nueve han aprobado todas las materias en los primeros 30 días de haberlas cursado, mientras que el 50% de los alumnos con promedio 5 terminan de aprobar todas las materias después del año de haberlas cursado. El valor  $p < 0.0001$  corresponde al `log_rank_test` que indica diferencias en las curvas para cada calificación promedio.



Se realizó una prueba chi-cuadrado para cuantificar la asociación entre el tiempo necesario para aprobar y la calificación ( $p < 0,001$ ). A medida que el tiempo para rendir aumenta, las calificaciones decrecen. Por último se utilizó una regresión logística para ajustar si el estudiante acreditó o no la asignatura la primera vez que se presentó en función del tiempo transcurrido entre que cursó la materia y la rindió por primera vez, más la variable asignatura, más la interacción de ambas. Este modelo saturado no dio diferencias significativas en la interacción, por lo que se ajustó el siguiente modelo para estimar probabilidades:

$$p(X) = \frac{e^{(\beta_0 + \beta_1 * tiempo + \sum_{i=2}^{14} \beta_i * m_i)}}{1 + e^{(\beta_0 + \beta_1 * tiempo + \sum_{i=2}^{14} \beta_i * m_i)}}$$

Donde  $m_i$  representa la  $i$  – esima materia

En este caso, tanto el tiempo como el efecto de las asignaturas fueron significativas. A continuación se muestran las probabilidades de aprobar para cada materia en diferentes tiempos:

	Tiempo (días)					
	1	30	180	365	700	1000
<b>Materia 1</b>	0,901	0,897	0,872	0,836	0,748	0,648
<b>Materia 2</b>	0,642	0,631	0,573	0,500	0,369	0,266

<b>Materia 3</b>	0,670	0,659	0,604	0,531	0,399	0,291
<b>Materia 4</b>	0,677	0,667	0,611	0,539	0,406	0,298
<b>Materia 5</b>	0,683	0,672	0,618	0,546	0,413	0,303
<b>Materia 6</b>	0,565	0,554	0,494	0,420	0,298	0,208
<b>Materia 7</b>	0,489	0,478	0,419	0,349	0,239	0,162
<b>Materia 8</b>	0,485	0,474	0,415	0,345	0,236	0,160
<b>Materia 9</b>	0,552	0,540	0,480	0,408	0,287	0,199
<b>Materia 10</b>	0,804	0,796	0,755	0,696	0,572	0,453
<b>Materia 11</b>	0,856	0,850	0,816	0,768	0,659	0,545
<b>Materia 12</b>	0,666	0,655	0,599	0,527	0,394	0,287
<b>Materia 13</b>	0,752	0,743	0,695	0,629	0,498	0,380

El paso de la educación del nivel medio al universitario no es simple, y la posibilidad de transitar de forma exitosa los primeros años depende de múltiples causas. En este trabajo nos hemos concentrado en aquellas que pueden ajustarse desde la administración académica de la institución.

Los resultados obtenidos permiten observar una asociación entre el tiempo que demoran los estudiantes en rendir y sus calificaciones. Para poder avanzar en la determinación de una relación de tipo causa-efecto, se planea realizar una experiencia que permita controlar factores adicionales que intervienen en la calificación de los estudiantes (factores socio-económicos, calificación promedio durante el cursado, promedio colegio secundario, etc). Por otra parte, este estudio permitirá instalar la discusión sobre la conveniencia del régimen de promoción que hoy cuentan algunas materias de la carrera.

## REFERENCIAS

- [1] Garbanzo Vargas, Guiselle María (2007). "Factores asociados al rendimiento académico en estudiantes universitarios, una reflexión desde la calidad de la educación superior pública". *Revista Educación*, 31(1): 43-63.
- [2] Kalbfleisch, J. y Prentice, R. (2002). *The Statistical Analysis of Failure Time Data*. Second Edition. United States of America.
- [3] Therneau, T. (2020). *A Package for Survival Analysis in R*. R package version 3.2-7, <https://CRAN.R-project.org/package=survival>.
- [4] Kassambara, A., Kosinski, M. y Bieček, P. (2020). *survminer: Drawing Survival Curves using 'ggplot2'*. R package versión 0.4.8, <https://CRAN.R-project.org/package=survminer>.

## **COMPETENCIAS DOCENTES EN DOCTORES EN CIENCIAS BIOLÓGICAS, ANÁLISIS DESDE SUS PERCEPCIONES**

SIMONIELLO M. FERNANDA<sup>1</sup>, ODETTI HECTOR<sup>1</sup>, TEMPORETTI FELIX<sup>2</sup>, AVILA OLGA<sup>1</sup>,  
CONTINI LILIANA<sup>1</sup>

<sup>1</sup> *Facultad de Bioquímica y Ciencias Biológicas, Universidad Nacional del Litoral (UNL), Santa Fe, Argentina*

<sup>2</sup> *Facultad de Psicología, Universidad Nacional de Rosario (UNR), Rosario, Argentina  
olga.beatriz.avila@gmail.com*

### **RESUMEN**

El objetivo de este trabajo es describir las concepciones referidas a la docencia universitaria y actividades relacionadas que poseen los jóvenes docentes, doctores en Ciencias Biológicas identificando cuáles podrían direccionar su proceso formativo como docentes universitarios. Se realizaron entrevistas en las que los docentes debían describir cada una de las actividades que realizaban, sus preferencias y sus tensiones. Se realizó un análisis de correspondencias múltiples con las categorizaciones (Cerca, Medianamente Cerca y Lejos) que los entrevistados hacían de esas tareas relacionadas con su accionar docente. Los resultados permitieron observar que es posible asociar que el Profesor se percibe cerca de la evaluación y planificación de la clase mientras que el auxiliar docente tiene una percepción de su “Yo docente” medianamente cerca de la organización de la clase. Por otra parte, los auxiliares se perciben cerca de la comunicación con los alumnos y los profesores medianamente cerca de las TIC. Mientras que Investigar sobre la enseñanza, en particular en la docencia universitaria en Ciencias Biológicas, se han convertido en un reducto especializado casi exclusivo de pedagogos ya que son escasos los trabajos multidisciplinarios.

**Palabras clave:** *educación superior, concepciones, análisis de correspondencias múltiples.*

### **INTRODUCCIÓN**

El profesorado universitario en Ciencias Experimentales basa su enfoque principalmente en la disciplina, pero, además, a estos profesores se les plantea la necesidad de una más clara profesionalización en el conjunto de competencias docentes que los acreditan como formadores bien formados. En este contexto, los docentes-investigadores realizan multiplicidad de tareas: la enseñanza de la asignatura de grado, la formación de alumnos en investigación disciplinar, las publicaciones, la administración de subsidios, e incluso actividades de gestión y extensión. Este trabajo toma como punto de partida la Universidad Nacional del Litoral (UNL), que nació luego de la Reforma de 1918 y dio origen a la formación de la Universidad Nacional de Rosario (UNR) y posteriormente a la Universidad Nacional de Entre Ríos (UNER), situadas en la zona Litoral de Argentina.

El objetivo de este trabajo es describir las concepciones referidas a la docencia universitaria y actividades relacionadas que poseen los jóvenes docentes, doctores en Ciencias Biológicas identificando cuáles podrían direccionar su proceso formativo como docentes universitarios.

## MATERIALES Y MÉTODOS

Se realizaron entrevistas, entre 2019 y 2020, a doctores en Ciencias Biológicas, docentes universitarios de UNL, UNR y UNER, de aproximadamente 40 años o menos, que aceptaron participar. El 50% de las entrevistas se realizó de manera virtual utilizando como mediadora la plataforma Zoom debido al contexto actual de pandemia. Se les presentó a los entrevistados un gráfico que incluía nueve cuadros con actividades que suelen realizar los docentes universitarios (Seleccionar los contenidos, Comunicarse con los alumnos, Manejar las TIC, Organizar las clases, Tutorías o consultas, Evaluar, Trabajar en equipo, Planificar, Investigar), un círculo que representaría a las que consideran tareas básicas correspondientes a la docencia universitaria en general y en su centro un círculo relacionado a su accionar docente (Figura 1). Se les solicitó que acerquen los cuadros de texto con las actividades hacia su “Yo docente” en función de la importancia o de la frecuencia con que realizan dichas tareas. Posteriormente debían describir cada una de ellas, sus preferencias y sus tensiones. Para los docentes que realizaron la entrevista de forma presencial la imagen estaba montada en una superficie metálica y las tareas tenían imanes, para las entrevistas virtuales se utilizó un archivo editable con extensión ppt. En cada una de estas actividades las posibles respuestas eran: Cerca (C), Medianamente Cerca (M) y Lejos (L).

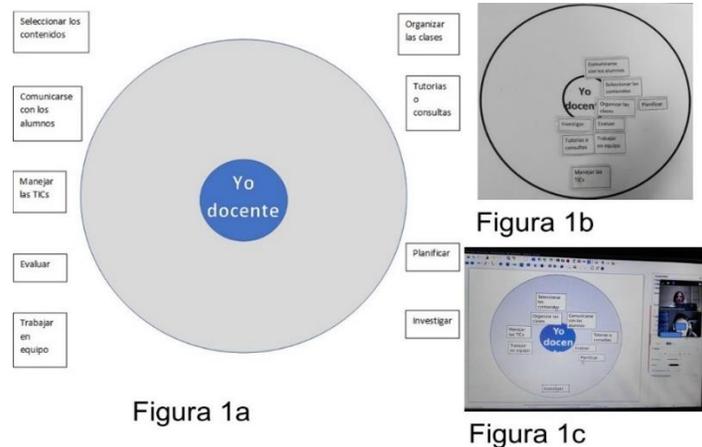


Figura 1. Gráfico utilizado para realizar las entrevistas. 1a: actividades propuestas para analizar. 1b: dispositivo inmantado. 1c: archivo editable para entrevista virtual

**Análisis estadístico:** Se realizó un análisis descriptivo de las variables edad, sexo, cargo docente y universidad. Se recategorizó la variable cargo docente en: Profesores (Titular y Adjunto) y Auxiliares (Ayudantes y Jefes de Trabajos Prácticos). Posteriormente se realizó un análisis de correspondencias múltiples [1].

## RESULTADOS Y DISCUSIÓN

El análisis de los resultados se sitúa en y desde el campo teórico educativo utilizando múltiples disciplinas para comprender y analizar las concepciones de los docentes universitarios noveles, doctores en Ciencias Biológicas. Cuando hablamos de “concepciones”, hacemos referencia a las percepciones, ideas o definiciones construidas por los sujetos en su interacción con otros, en el marco de su ámbito cotidiano. Estas actúan de filtro y elemento decodificador de las informaciones procedentes de otros ámbitos de investigación. Por ello, consideramos que tales concepciones constituyen un eje transversal de la evolución profesional del profesor.

En las entrevistas participaron un total de 27 docentes, 8 varones y 19 mujeres, con un promedio de edad de 36 años de los cuales 21 tenían cargo de auxiliar y 6 cargo de profesor. Las entrevistas duraron entre 30 y 45 minutos, no mostrando diferencias entre sus respuestas cuando se consideró la metodología de la entrevista (presencial o virtual). Todos respondieron que elegían la frecuencia con la que realizaban las tareas para acercarlas a su “yo docente”. Para el análisis de correspondencias múltiples, inicialmente se consideraron todas las variables relevadas de manera conjunta. Con este primer análisis no se obtuvo una explicación significativa desde la disciplina. Se procedió entonces a considerar la naturaleza de las variables y se formaron dos grupos, uno relacionado con las actividades docentes propiamente dichas (G1) y el otro relacionado con actividades de comunicación e investigación (G2). El análisis de correspondencias con las variables: Organización, Planificación, Evaluación y cargo docente (G1), permitió observar que es posible asociar que el Profesor se percibe cerca de la evaluación y planificación de la clase mientras que el auxiliar docente tiene una percepción de su “Yo docente” medianamente cerca de la organización de la clase como tal (Figura 2). La inercia observada para este análisis fue de 52%

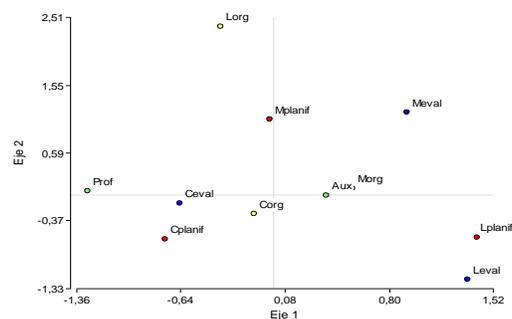


Figura 2: Biplot con las variables Cargo docente, Planificación, Evaluación y Organización de las clases

Según Zabalza [2], organizar las clases incluye la organización de los espacios, la selección de la metodología y como abordar los contenidos. Mientras que Planificar en el proceso de enseñanza-aprendizaje incluye: diseñar programas adaptados a las circunstancias, seleccionar los distintos dispositivos, comunicar los contenidos y facilitar el aprendizaje. Con respecto a las evaluaciones, observamos en los docentes dos situaciones enfrentadas. Por un lado, para los auxiliares insatisfacción por la actividad, ya que su preferencia está en dar clase y abrir nuevas perspectivas científicas, culturales y profesionales. Por otro lado, están los profesores, para quienes enseñanza y evaluación están intrincadamente relacionadas.

Para el análisis de correspondencias del G2 la inercia explicada fue inferior al 50 % y en él se observa que los auxiliares se perciben cerca de la comunicación con los alumnos y los profesores medianamente cerca de las TIC. Una de las actividades en la que se centran los auxiliares es en la Comunicación con los alumnos, que se constituye como la primera vía de transmisión educativa. Mientras que la utilización de las TIC es hoy una herramienta fundamental para el ejercicio de la docencia universitaria, por lo general, conduce a tensiones vinculadas con la dificultad entre elegir metodologías sincrónicas o asincrónicas, ya que las primeras potencian las diferencias de conectividad y acceso entre los estudiantes, y en las segundas los estudiantes pierden la contención y el acompañamiento, ya que el aprendizaje se constituye como un proceso interactivo. Con relación a Investigar sobre la enseñanza, en



particular en la docencia universitaria en ciencias biológicas, se ha convertido en un reducto especializado casi exclusivo de pedagogos ya que son escasos los trabajos multidisciplinarios.

## **REFERENCIAS**

- [1] Greenacre, M. J. 1984, Theory and applications of correpondence analysis. Academic Press London.
- [2] Zabalza, M. Á. (2006). Competencias docentes del profesorado universitario: calidad y desarrollo profesional. Narcea Ediciones. Madrid.

## ANÁLISIS DE LAS CARACTERÍSTICAS HÍDRICAS Y CLIMÁTICAS EN LA CUENCA DEL RÍO COLLÓN CURÁ, MEDIANTE MODELOS LINEALES GENERALIZADOS

ULLOA DIEGO<sup>1,4</sup>, HAIQUE ANA KARINA<sup>2,4</sup>, CARRASCO GABRIELA<sup>3,4</sup>, RABY GABRIEL<sup>4</sup>

<sup>1</sup>AIC - Autoridad Interjurisdiccional de las Cuencas de los ríos Limay, Neuquén y Negro  
Secretaría de Operación y Fiscalización

<sup>2</sup>Departamento de Estadística / Facultad de Economía / Universidad Nacional del Comahue

<sup>3</sup>Centro Universitario de Estudios en Salud, Economía y Bienestar, Universidad Nacional del Comahue

<sup>4</sup>Maestría en Estadística Aplicada, Universidad Nacional del Comahue  
dulloo@aic.gob.ar

### RESUMEN

La cuenca del río Collón Curá, ubicada en la zona norte de la Patagonia, despierta un interés particular ya que convergen sobre la misma diversos usos del recurso hídrico tales como: la generación hidroeléctrica para el Sistema Interconectado Nacional, el abastecimiento de agua para el desarrollo de economías de subsistencia local, la actividad frutícola tanto para el mercado local como el de exportación, el abastecimiento de agua para el uso residencial, la industria petrolera, y la minería. Sin embargo, no hay muchos estudios, desde el punto de vista climático que permitan describirla. Con el objetivo de caracterizar dicha cuenca, se seleccionaron 30 años correspondientes al período entre 1990 y 2019, clasificándolos en tres categorías: Seco, Medio, Húmedo. Dicha clasificación se realizó con los datos hidrológicos de cada año correspondiente a los derrames en el Collón Curá. Adoptando como variable de respuesta la categorización mencionada. Se plantearon dos posibles modelos donde se tomaron como variable regresoras tres índices climatológicos, en sus versiones anuales y estacional (abril-agosto): índice de oscilación meridional (SOI), índice Niño Oceánico (ONI) y índice de oscilación antártico (AAO). El primer modelo lineal generalizado propuesto fue un modelo logit con categoría de referencia y el segundo modelo fue un modelo odd proporcional. En ambos casos se aplicaron técnicas stepwise en búsqueda del modelo más parsimonioso. En base al AIC de ambos modelos y por la simplicidad en la interpretación, se consideró mejor la primera opción. La forzante climática de variabilidad interanual más importante en la clasificación de años fue el índice SOI estacional para el período abril-agosto. Para ratificar el modelo elegido, se tomó la información de los años 1985 a 1989 para hacer predicciones. Para el período de 5 años, el modelo ajustó bien los años secos y medios, pero no así los años húmedos.

**Palabras clave:** *modelo logit con categoría de referencia, modelo odd proporcional, variabilidad hidrológica interanual, Índices climáticos, Cuenca Collón Curá.*

### INTRODUCCIÓN

En la actualidad, existen dificultades para explicar el comportamiento hidrológico de la cuenca del Collón Curá a partir de variables climatológicas del Pacífico. El monitoreo y eventual pronóstico de variables hidrometeorológicas podría ayudar a las diferentes actividades económicas de la región, identificando patrones comunes previos y simultáneos observados en

los campos de temperatura de la superficie del mar, de circulación hemisférica (geopotencial y viento en diferentes niveles) y regional (viento en capas bajas, disponibilidad de humedad) para los años secos, medios y húmedos [2]. Desde el punto de vista agrometeorológico la predicción anticipada de un evento extremo húmedo o seco permitiría un manejo más eficiente del riego y de los cultivos en general [1].

El objetivo de este trabajo fue realizar un análisis de la cuenca del Collón Curá relacionando los años, del período bajo análisis, clasificados en seco, medio y húmedo con datos climáticos de la zona del Pacífico: SOI, ONI y AAO.

## MATERIALES Y MÉTODOS

La variable de respuesta se obtuvo de clasificar el nivel de precipitaciones de cada año del período (1990 – 2019) analizado en función de la de excedencia de los derrames acumulados para la serie de datos desde 1941 hasta 2020<sup>13</sup>. Se optó como criterio de clasificación los cuartiles: SECO ( $\leq 25\%$ ), MEDIO ( $>25\%$  y  $<75\%$ ) y HÚMEDO ( $\geq 75\%$ ) [2].

Inicialmente, se planteó un modelo logit con categoría de referencia [4] (seco) con la totalidad de las variables regresoras. Se recurrió a la función stepAIC del módulo MASS, aplicada en la dirección 'backward' hasta encontrar un modelo con menos variables. Para optar por la versión completa o más parsimoniosa del modelo, se recurrió a una prueba de cociente de verosimilitud [4]. En segundo lugar, se planteó un modelo odd proporcional [4] completo esto es con las seis variables regresoras y se realizó una prueba de proporcionalidad. Nuevamente se realizó un proceso Stepwise para buscar un modelo con menos variables y para optar por la versión completa o más parsimoniosa se aplicó una prueba de cociente de verosimilitud. Para optar por uno de los dos modelos resultantes, se compararon sus respectivos valores de AIC.

## RESULTADOS Y DISCUSIÓN

Para ensayar una selección entre ambos modelos, se tomó en consideración el criterio de AIC: El AIC del modelo final logit con categoría de referencia para variable tratada como nominal, fue de 55,65, mientras que el AIC del modelo final Odds proporcional para variable tratada como ordinal, fue de 53,44. Al no resultar concluyentes los valores de AIC, se consideró mejor opción, por su simplicidad el modelo logit.

Modelo:

$$\text{Logit}[\pi_{\text{Medio}}] = \log\left(\frac{\pi_{\text{Medio}}}{\pi_{\text{Seco}}}\right) = 0.7102445 - 0.1763670 * \text{SOI}_{\text{abr\_ago}}$$

$$\text{Logit}[\pi_{\text{Húmedo}}] = \log\left(\frac{\pi_{\text{Húmedo}}}{\pi_{\text{Seco}}}\right) = -1.4798128 - 0.3276094 * \text{SOI}_{\text{abr\_ago}}$$

Como se puede apreciar, la forzante climática de variabilidad interanual más importante en la clasificación de años fue el índice SOI estacional para el período abril-agosto.

Al aplicar función exponencial a las pendientes de las funciones de regresión, se puede hacer la siguiente interpretación: Primero, a medida que el índice SOI<sub>abr\_ago</sub> aumenta en una unidad, las chances de observar un año con nivel medio, es 0,83 veces respecto a observar un año seco. Es decir, a medida que aumenta el índice SOI<sub>abr\_ago</sub> la probabilidad de obtener un año medio es menor que la probabilidad de obtener uno seco, un 17% menor. Segundo, a medida que el índice SOI<sub>abr\_ago</sub> aumenta en una unidad, las chances de observar un año con nivel húmedo, es 0,72 veces respecto a observar un año seco. Es decir, a medida que aumenta el

<sup>13</sup> Datos proporcionados por la Autoridad Interjurisdiccional de Cuenas

índice SOI\_abr\_ago la probabilidad de obtener un año húmedo es menor que la probabilidad de obtener uno seco, un 28% menor.

Se realizó un ejercicio para evaluar las virtudes del modelo con los datos del período 1985-1989:

Para el período de 5 años de validación se puede observar que el modelo ajusta bien los años secos y medios. No se pudo predecir el año húmedo. Los resultados obtenidos sugieren una sobreparametrización del modelo, para futuros estudios podría considerarse un análisis con variable de respuesta de dos categorías (Seco, Húmedo) criterio empleado por algunos estudios usados como referencia en este trabajo. Otra variante que amerita ser analizada, es el cambio de criterio en la determinación de la variable de respuesta hidrológica (clasificación de años). El método de clasificación hidrológica anual debería considerar las precipitaciones acumuladas anuales además de los derrames acumulados salientes de la cuenca.

Tabla 1: Resultados obtenidos de aplicar el modelo a los años correspondientes al período 1985-1989

Datos <sup>14</sup>			Predicción del modelo		
Año	Clase	SOI_abr_ago	Seco	Medio	Húmedo
1985	Medio	2,8	42,9%	53,2%	3,9%
1986	Húmedo	-0,02	30,6%	62,4%	7,0%
1987	Medio	-19,74	0,5%	31,0%	68,6%
1988	Seco	6,2	58,4%	39,8%	1,7%
1989	Seco	9,2	70,8%	28,4%	0,8%

## REFERENCIAS

- [1] Romero P., Garbarini E., Gonzáles M. (2014) “Características hídricas y climáticas del norte Patagónico”. Instituto Nacional del Agua - IFRH 2er Encuentro de Investigadores en Formación en Recursos Hídricos. Argentina. <https://www.ina.gov.ar/ifrh-2014/Eje3/3.33.pdf>
- [2] Romero P., Garbarini E., Gonzáles M. (2016) “Patrones de temperatura de mar y circulación atmosférica asociados a eventos húmedos y secos en el Comahue”. Instituto Nacional del Agua - IFRH 3er Encuentro de Investigadores en Formación en Recursos Hídricos. Argentina. <https://www.ina.gov.ar/ifrh-2018/index.php>
- [3] Groch, D., Cogliati, M.G., Finessi, F.G. (2020). “Influencia de ENOS en la hidrometeorología de la cuenca alta del río Neuquén”. Geograficando, vol.(16), núm. 1, 2020.
- [4] Faraway, J. (2016) “Extending the Linear Model with R. Generalized Linear, Mixed Effects and Nonparametric Regression Models” Second Edition, FL, USA.

<sup>14</sup> Datos proporcionados por la Asociación Interjurisdiccional de Cuencas

## ELIMINACIÓN DE NUTRIENTES Y MATERIA ORGÁNICA EN AGUA RESIDUAL URBANA POR HUMEDALES CONSTRUIDOS

VAIRA STELLA<sup>1</sup>, CARRIO JOSEFINA<sup>1</sup>, PIZARRO ANA<sup>2</sup>, ZERBATO MARIEL<sup>2</sup>, MODINI LAURA<sup>2</sup>

<sup>1</sup>Departamento de Matemática de la FBCB – LISEB / Universidad Nacional del Litoral

<sup>2</sup>Cátedra de Tratamiento de Efluentes de la FBCB / Universidad Nacional del Litoral  
svaira@fbc.unl.edu.ar

### RESUMEN

Este estudio investiga la eficiencia de humedales contruidos de flujo subsuperficial vertical (HFVSS), con arena y coque como sustrato, para eliminar nutrientes y materia orgánica de agua residual urbana. Se construyeron 3 sistemas con plantas (humedales) y uno sin planta. Los mismos fueron alimentados diariamente con efluente cloacal real, en forma intermitente y se dejaban drenar libremente hasta que el nivel del líquido descendía 15 cm por debajo de la superficie (HFVSS parcialmente saturado). Se analizaron los siguientes parámetros de contaminación: Demanda química de oxígeno, amonio, nitrito, nitrato, Fósforo reactivo total y soluble. Para el análisis estadístico se aplicaron varios modelos lineales, de acuerdo al objetivo de investigación y al tipo de variable estudiada. Los 4 sistemas lograron remover DQO y amonio con eficiencias promedio superior al 58,3% y 23,5%, pero con diferencias significativas entre los sistemas ( $p < 0,05$  para ambos parámetros). Además, los humedales retuvieron fosforo a diferencia del sistema sin plantas, que presentó liberación con diferencias de retención altamente significativas ( $p < 10^{-3}$ ).

**Palabras clave:** *humedal, cloacal, tratamiento, nutrientes, materia orgánica, modelos lineales robustos.*

### INTRODUCCIÓN

Los humedales artificiales son sistemas diseñados y contruidos para simular la transición entre el ambiente terrestre y el acuático. Su propósito es aprovechar los procesos fisicoquímicos y biológicos que ocurren al unir sustratos del suelo, plantas y microorganismos para la degradación de materia orgánica y nutrientes de las aguas residuales (Arias y col., 2010). Debido al bajo costo de instalación, su simplicidad de operación y mantenimiento y muy bajo requerimiento energético, representan una alternativa atractiva para el tratamiento de agua residual en pequeñas comunidades urbanas (Romero Aguilar y col., 2009; Vymazal, 2014). Sin embargo, la eficiencia de remoción de contaminantes depende de diversos factores, como el diseño, los materiales del lecho y el régimen de alimentación. En este trabajo se propone evaluar la eficiencia de remoción de nutrientes y materia orgánica en agua residual urbana por humedales contruidos de flujo subsuperficial vertical parcialmente saturados empleando coque y arena como sustratos. (Proyecto: ASaCTel IO-2018-00086 *Humedales electroquímicos para el tratamiento sustentable de agua residual urbana en pequeñas comunas de la provincia de Santa Fe*. Directora: Laura Modini)

### MATERIALES Y MÉTODOS

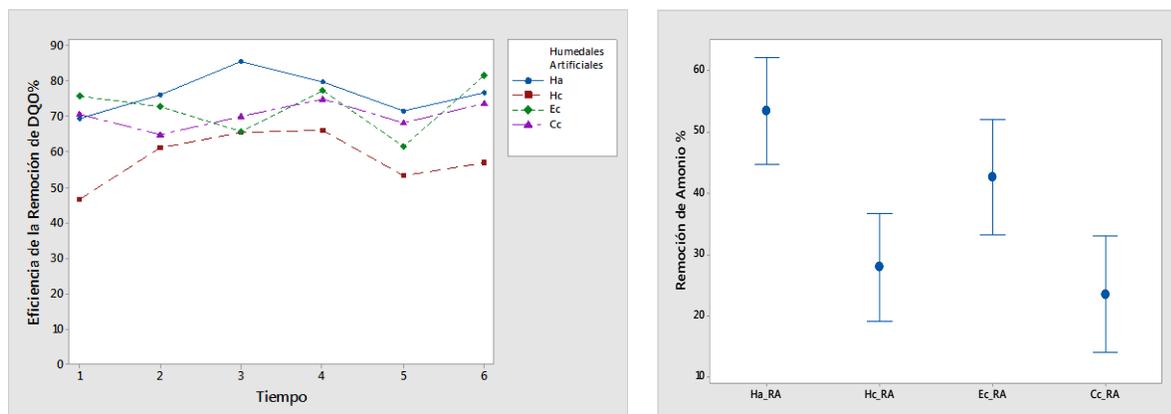
Se construyeron 3 sistemas de humedales a escala micropiloto, de tipo subsuperficial vertical descendente. Para esto se utilizaron tubos de PVC de 20 cm de diámetro y 70 cm de profundidad, con un falso fondo en la base. Como sustratos se usó arena y coque. Las configuraciones evaluadas fueron: Ha (55 cm de arena gruesa), Hc (55 cm de coque), Ec (de abajo hacia arriba: 20 cm de arena, 10 cm de coque, 20 cm de arena y 5 cm de coque). En cada humedal se plantaron ejemplares de *Schoenoplectus californicus* a escasa profundidad, de manera que las raíces permanecieran principalmente en la parte superior del reactor. Para evaluar el aporte de las plantas a la depuración, se construyó un cuarto sistema Cc, similar a Ec, pero sin plantas. Se usó efluente cloacal crudo provisto por la planta depuradora de líquidos cloacales de la ciudad de Santo Tomé (Santa Fe, Argentina), previamente sedimentado durante 2 h. Los sistemas fueron alimentados diariamente, con un único pulso, y se dejaban drenar libremente hasta que el nivel del líquido descendía 15 cm por debajo de la superficie. El tiempo de residencia hidráulica (TRH) fue 24 h. Para evaluar la eficiencia de remoción, semanalmente se tomaron muestras a la entrada (líquido cloacal) y a la salida (efluente tratado) de cada sistema y se analizaron los siguientes parámetros: Demanda Química de Oxígeno (DQO), amonio, nitrito, nitrato, Fósforo Reactivo Total (PRT) y Fósforo Reactivo Soluble (PRS). Todas las técnicas se realizaron de acuerdo con *Standart Methods* (APHA, AWWA, WEF, 2012).

#### *Análisis estadístico*

El tratamiento estadístico estuvo sujeto a la naturaleza de las variables y al objetivo de la investigación. Se realizó Regresión Lineal (RL) y Análisis de la Variancia (ANOVA) desde el paquete `lm()` y `aov()` del software R. Además, cuando fue necesario, se aplicaron modelos lineales robustos para tamaños de muestras pequeños, ya que éstos ponderan la influencia de los casos atípicos, produciendo coeficientes y errores estándar más confiables para mejorar de esta forma las estimaciones. En este caso se utilizó el paquete `MASS()` (Venables and Ripley, 2002). Nivel de significancia  $\alpha$  prefijado de 0,05.

## RESULTADOS Y DISCUSIÓN

El rendimiento de los cuatro humedales artificiales, en cuanto a la remoción de materia orgánica (DQO%), se muestra en la Figura 1 A. Los sistemas Ha, Ec y Cc eliminaron DQO con eficiencias similares ( $76,6 \pm 5,8\%$ ;  $72,5 \pm 7,5\%$  y  $70,4 \pm 3,4\%$ , respectivamente). En Hc la eficiencia fue inferior a la de Cc ( $p = 0,010$ ), con un porcentaje medio de 58,3%; probablemente debido a que la carga aplicada en este sistema (promedio de 53 gDQO/m<sup>2</sup>.d) prácticamente duplicó la que recibió el resto de los reactores (promedio de 28,8 gDQO/m<sup>2</sup>.d HA y 37,1 gDQO/m<sup>2</sup>.d Ec y Cc). Se desprende claramente de la Figura 1 A que los cuatro sistemas proporcionaron un tratamiento estable, sin hallarse un modelo en el tiempo y pudieron eliminar DQO del efluente cloacal previsiblemente, aunque el sistema Hc con una eficiencia de tratamiento ligeramente menor.



A

B

**Figura 1.** A. Eficiencia de la remoción de DQO%. B. Eficiencia de la remoción de Amonio (%). Ambos parámetros en los cuatro sistemas de tratamiento durante las 6 semanas de ensayo.

Para el caso del FRT (mg/L) en el efluente tratado por los 4 sistemas, se hallaron diferencias significativas (valor p asociado al Test F de Welch  $< 10^{-3}$  entre los 4 sistemas). El menor IC95% correspondió a Ha [2,5; 3,9] contra Cc que presentó los siguientes valores estimados [5,7; 7,5]. Un comportamiento similar se encontró en el FRS efluente (mg/L). La mayor concentración de fósforo a la salida de Cc se podría atribuir a la ausencia de plantas en este sistema. El valor de  $R^2$  ajustado para las observaciones fue de 76,5% y 85,4% para cada variable respuesta FRT y FRS, con un valor medio en el cloacal crudo de FRT de 5,9 mg/L y de FRS de 5,3 mg/L.

Por otro lado, hay homogeneidad en el comportamiento en los 4 sistemas cuando se analiza las concentraciones (mg/L) de nitrito y nitrato en el efluente tratado. Pero en cuanto a la remoción de Amonio (%), se hallaron diferencias significativas (valor  $p < 10^{-3}$  del ANOVA). Ec y Ha resultaron más eficientes para eliminar amonio (Figura 1.B), siendo la Remoción porcentual de Amonio en Ec de [33,3; 52,1], incluyendo un valor individual de 57,8%.

En este trabajo se analizó la eficiencia de humedales de tratamiento para descontaminar agua residual urbana. Las remociones porcentuales de DQO en los humedales diseñados pueden ser revisadas, ya que Hc recibió una carga orgánica más elevada debido a la mayor porosidad del lecho de coque con respecto al de arena. El tipo de operación propuesta (alimentación intermitente con substrato parcialmente saturado), permitió remover efectivamente nitrógeno del efluente cloacal, principalmente por Ha y Ec. Asimismo, los humedales presentaron retención de fósforo, a diferencia del sistema Cc que no tenía plantas.

## REFERENCIAS

- [1] Rice, E. W.; Baird, R. B.; Eaton, A. D.; Clesceri, L. S. (2012). *Standard methods for the examination of water and wastewater*. 22nd ed. American Public Health Association, USA.
- [2] Arias, S.; Betancour, F.; Rojas, G.; Giraldo, J. y Hernández, M. (2010). "Fitorremediación con humedales artificiales para el tratamiento de aguas residuales porcinas". *RIT* 74:12-22.
- [3] Liam, D.; Yaquan, Z.; Xiaohong, Z.; Wenke, W. (2015). "Nutrient and organics removal from swine slurry with simultaneous electricity generation in an alum sludge-based constructed wetland incorporating microbial fuel cell technology". *Chemical Engineering Journal* 266:74-81.
- [4] R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.



- [5] Romero Aguilar, M.; Colin Cruz, A.; Salinas Sánchez, E. y Hernández Ortiz, M. (2009). "Tratamiento de aguas residuales por un sistema piloto de humedales artificiales: Evolución de la remoción de la carga orgánica". *Rev. Int. Contam. Ambient.* 25(3): 157-167.
- [6] Shibao, L.; Xiaoling, Z.; Jianhua, W.; Liang, P. (2016). "Impacts of different media on constructed wetlands for rural household sewage treatment". *Journal of Cleaner Production* 127: 325-30.
- [7] Venables, W. N., Ripley B. D. (2002). *Modern Applied Statistics with S*, Fourth edition. Springer, USA.
- [8] Vymazal, J. (2014). "Constructed wetlands for treatment of industrial wastewaters: A review". *Ecol. Eng.* 73: 724-751.

## BÚSQUEDA DE ESTRUCTURA GENÉTICA: COMPARACIÓN DE ALGORITMOS

VIDELA M EUGENIA<sup>1,2,3</sup>, BRUNO CECILIA<sup>1,2</sup>

<sup>1</sup>Consejo Nacional de Investigaciones Científicas y Tecnológicas (CONICET)

<sup>2</sup>Cátedra de Estadística y Biometría. Facultad de Ciencias Agropecuarias (FCA). Universidad Nacional de Córdoba. Grupo asociado con UFyMA (INTA -CONICET)

<sup>3</sup>Universidad Nacional de Villa María, Argentina.

eugeniavidela12@gmail.com

### RESUMEN

Existe una amplia disponibilidad de métodos de clasificación y ordenamiento para identificar grupos de individuos caracterizados mediante técnicas multivariadas como los marcadores del tipo Single Nucleotide Polymorphism (SNP). Este tipo de marcadores generan bases de datos multivariadas de alta dimensionalidad. El objetivo de este trabajo es comparar tres métodos de agrupamiento (jerárquico, no jerárquico y bayesiano) y un método de clasificación (basado en aprendizaje de máquina) en una base de datos molecular masiva para la búsqueda de estructura genética poblacional. La implementación de los métodos se realiza sobre una base de datos pública de maíz que cuenta con 942 líneas genotipadas mediante 899784 marcadores SNPs, donde identificaron 11 conglomerados o sub-poblaciones clasificadas con el software Admixture 1.23 que verificaron desde el conocimiento biológico de investigaciones genéticas en maíz y fue considerada como agrupamiento de referencia (“gold estándar”). En este trabajo se estimó la tasa de error de clasificación de los cuatro métodos de agrupamiento. El método bayesiano mostró mejor desempeño para clasificar los genotipos con una tasa de error del 18%, mientras que UPGMA y SVM fueron quienes más errores presentó en el agrupamiento de los individuos (58% y 60%, respectivamente).

**Palabras clave:** *estructura genética poblacional, algoritmos de agrupamiento, algoritmos de clasificación, clasificación supervisada, clasificación no supervisada, ordenamiento, análisis multivariado.*

### INTRODUCCIÓN

La creciente disponibilidad de datos provenientes de diferentes tipos de marcadores moleculares permite una caracterización exhaustiva de la diversidad genética en diferentes especies. Dado un conjunto de individuos caracterizados molecularmente, se espera que aquellos que presentan mayor similitud en su perfil genético, se encuentren relacionados, en algún grado de parentesco, y por lo tanto puedan agruparse definiendo poblaciones o grupos genéticos. Los marcadores moleculares han sido utilizados para estimar relaciones entre diversas líneas. En la actualidad, los marcadores moleculares de tipo *Single Nucleotide Polymorphism* (SNP), han cobrado importancia por la gran cantidad de información que proporcionan. La búsqueda de estructura genética poblacional (PGS) en una colección de datos de alta dimensión implica un incremento en la complejidad del manejo de bases de datos masivas como las generadas por los marcadores moleculares de tipo SNP. Diferentes métodos multivariados para identificar grupos han surgido como una alternativa para analizar con precisión el volumen masivo de datos generados por las aplicaciones modernas. El objetivo de

este trabajo es comparar tres métodos de agrupamiento (jerárquico, no jerárquico y bayesiano) y un método de clasificación (basado en aprendizaje de máquina) en una base de datos molecular masiva para la búsqueda de estructura genética poblacional.

## MATERIALES Y MÉTODOS

El conjunto de datos consiste en 942 líneas de maíz genotipadas mediante 899784 marcadores SNPs derivados de RNA-Seq publicados por Mazaheri *et al.* (2019) donde se identificaron diez conglomerados o sub-poblaciones más una subpoblación mezcla con el software Admixture 1.23 (Alexander, 2009) y verificadas desde el conocimiento biológico de investigaciones genéticas de maíz. De acuerdo con esta clasificación, el panel se dividió en cuatro subpoblaciones stiff stalk (SS), dos subpoblaciones non-stiff stalk (NSS), una subpoblación de líneas públicas de origen amplio, una Iodent (IDT), una subpoblación de sweet corn, una de popcorn y una subpoblación tropical inbreds. Un total de 201 individuos endogámicos con menos de 0.5 de probabilidad de pertenencia a cualquiera de las subpoblaciones antes mencionadas fue clasificado como un grupo "mixto". El conjunto de 11 grupos identificados fue usada en este trabajo como referencia o "gold standard". La divergencia genética promedio entre las subpoblaciones, en términos del estadístico  $F_{ST}$  introducido por Wright (1949), fue de  $0.0239 \pm 0.009$ . Las poblaciones menos divergentes fueron Sweet corn y las líneas públicas de origen amplio ( $F_{ST} = 0.007$ ) y las poblaciones más diferenciadas fueron SS-B73 y NSS-Mo17 con un valor de  $F_{ST}$  de 0.047.

Tres métodos de agrupamiento y un método de clasificación: el método jerárquico Unweighted Pair Group Method with Arithmetic Mean (UPGMA) (Sokal y Michener, 1958), el método no jerárquico k-means (MacQueen, 1967), el método bayesiano de Structure (Pritchard *et al.*, 2000) y el método de clasificación basado en aprendizaje de máquinas Support Vector Machine (SVM) (Cortes y Vapnik, 1995) fueron comparados respecto a su capacidad de detectar los agrupamientos propuestos por Mazaheri *et al.* (2019). Para ello se evaluó el agrupamiento/clasificación de los cuatro métodos en la implementación desde dos a  $k=11$  conglomerados. Para el método SVM se utilizó el vector de clasificación obtenido por Structure como clasificación inicial de aprendizaje. Para evaluar el desempeño de cada método se calculó el porcentaje de individuos mal clasificados según la clasificación de referencia.

## RESULTADOS Y DISCUSIÓN

Los métodos UPGMA y SVM obtuvieron los mayores porcentajes de mala clasificación, es decir, agrupó individuos en subpoblaciones diferentes a las establecidas en el "gold standard", mientras que Structure fue el método que presentó menor error de asignación de individuos a subpoblaciones según las clasificaciones ya establecidas por Admixture en Mazaheri *et al.* (2019) (Tabla 1).

Tabla 1. Porcentaje de mala clasificación de cuatro métodos de agrupamiento sobre conjunto de datos reales publicados por Mazaheri *et al.* (2019)

Método	Porcentaje de Mala Clasificación
UPGMA	0.5816464
k-means	0.3130904
Structure	0.1835358
Support Vector Machine	0.6018893

Con Structure, se obtuvo una perfecta clasificación de las cuatro subpoblaciones stiff stalk (SS-B13, SS-B37, SSB73 y SS-BSSSC0), de la subpoblación tropical y de una de las subpoblaciones non-stiff stalk (NNS-Mo17). La otra subpoblación non-stiff stalk fueron bien clasificada en un 71,2%, confundiendo algunos individuos de esta población con sweet corn. En la subpoblación lodent el 97,1% de los individuos tenían una correctamente clasificación, mientras que el resto pertenecían a sweet corn. El 65,3% de la subpoblación de líneas públicas de origen amplio fue correctamente agrupada, confundiéndose el resto de las líneas con popcorn y líneas tropicales. La subpoblación de popcorn fue bien clasificada en un 58,3%, confundiéndose en mayor medida con líneas tropicales, sweet corn y NNS-Mo17 en menor medida. Por último, la población de sweet corn fue confundida casi en su totalidad con popcorn (Figura 2).

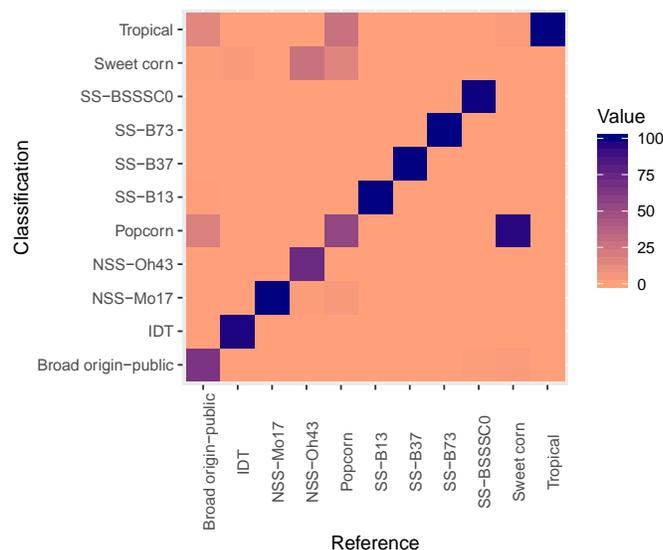


Figura 2. Heatmap de matriz de confusión de porcentaje de clasificación entre clasificación de referencia de conjunto de datos reales publicado por Mazaheri et al. (2019) y clasificación obtenida por el método Structure.

En el conjunto de datos públicos que se utilizó de ilustración se observó que el método Structure fue el de mejor desempeño para clasificar los genotipos, mientras que UPGMA y SVM fueron los que obtuvieron mayor porcentaje de mala clasificación.

## REFERENCIAS

- [1] Alexander DH, Novembre J. Fast Model-Based Estimation of Ancestry in Unrelated Individuals. *Genome research*. 2009;19:1655–64.  
Corina Cortes and Vladimir Vapnik. Support-vector network. *Machine Learning*, 20:273– 297, 1995.
- [2] MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of 5-th Berkeley*.



- [3] Mazaheri, M. et al. (2019), Data from: Genome-wide association analysis of stalk biomass and anatomical traits in maize, Dryad, Dataset, <https://doi.org/10.5061/dryad.n0m260p>
- [4] Mazaheri, M., Heckwolf, M., Vaillancourt, B., Gage, J. L., Burdo, B., Heckwolf, S., y Kaeppler, H. F. (2019). Genome-wide association analysis of stalk biomass and anatomical traits in maize. *BMC plant biology*, 19(1), 45.
- [5] Pritchard, J., M. Stephens and P. Donnelly (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155, 945 – 959.
- [6] Sokal, R. and Michener, C. (1958). A statistical methods for evaluating systematic relationships. *University of Kansas Science Symposium on Mathematical Statistics and Probability* 1, p. 17.
- [7] Wright, S. (1949). The genetical structure of populations. *Annals of eugenics*, 15(1), 323-354

## ANÁLISIS CANÓNICO DE CORRESPONDENCIAS Y ANÁLISIS DE COINERCIAS: UNA APLICACIÓN EN EL CAMPO DE LA ECOLOGÍA

ZULIANI ROXANNA<sup>1</sup>, LAFFITTE LORENA<sup>1,2</sup>, MUÑOZ SAAVEDRA JULIETA<sup>2</sup>, LAVALLE ANDREA<sup>1</sup>

<sup>1</sup>Departamento de Estadística Universidad Nacional del Comahue. Neuquén

<sup>2</sup>Dirección General de Biología Acuática. Subsecretaría de Ambiente. Neuquén  
matiroc@live.com.ar

### RESUMEN

Numerosos estudios ecológicos se basan en el análisis conjunto de dos tablas de datos: una tabla de frecuencias o abundancia de especies en diversos sitios de muestreo; y por otro lado, una tabla de información de variables ambientales medidas sobre esos mismos sitios. Existen diferentes técnicas enfocadas a investigar las relaciones entre las especies y el hábitat circundante. El Análisis Canónico de Correspondencias (ACC) es uno de los métodos de ordenación con restricciones más ampliamente utilizado y de interés en la comunidad de ecólogos. Otro enfoque más general es el Análisis de Co-inercia (ACol). En el primero la ordenación de los objetos representa la estructura de los datos que maximiza la relación con una segunda matriz de variables predictoras. En cambio, el ACol describe la co-estructura entre las dos tablas de datos representando lo mejor posible los cuadrados de las covarianzas entre ellas. Como características comunes a ambos métodos se destaca que son de gradiente directo y que se basan en la descomposición espectral. En tanto que una de las principales diferencias entre estas dos técnicas, es que en el ACC las frecuencias se consideran variables de respuesta y las variables ambientales son las explicativas. En cambio, el ACol no tiene restricción en el rol de las diferentes tablas como variables de respuesta o predictoras. En este trabajo se presentan las semejanzas y diferencias entre ambos métodos y se muestran los resultados de su aplicación a un conjunto de datos de sitios de muestreo de aguas caracterizados por variables fisicoquímicas y frecuencias de macroinvertebrados. Se utiliza el entorno de programación y análisis de datos de R.

**Palabras clave:** *datos de conjuntos múltiples, análisis de gradiente, ordenación con restricciones.*

### INTRODUCCIÓN

La ordenación con restricciones (ordenación canónica) asocia dos o más conjuntos de datos en el proceso de ordenación siendo una generalización de los métodos de ordenación simple. La ordenación es forzada a representar la parte de la variabilidad de una tabla de variables de respuesta a estar máximamente relacionada a otra tabla constituida por un conjunto de variables explicativas. En la comunidad de ecólogos este tipo de ordenación es ampliamente usada, estableciendo relaciones entre las especies y los factores ambientales que describen los mismos sitios ecológicos por sus características ambientales. Al contar con dos tablas con variables de diferente naturaleza, se requieren técnicas que permitan trabajar con toda la información simultáneamente, tales como ACC y ACol [1,2,3]. A partir de una aplicación en el campo de la ecología se presentan las semejanzas y diferencias entre ambos métodos y se muestran los resultados.

## MATERIALES Y MÉTODOS

En este trabajo, los individuos son sitios de muestreo en diferentes cuerpos de agua de la Cuenca Lácar - Hua Hum. La información proviene de dos grupos de variables registradas sobre los mismos sitios. Por un lado, frecuencias de familias de macroinvertebrados bentónicos indicadores de calidad de agua (Y); por otro lado, características fisicoquímicas del agua (X). Por lo tanto se tienen dos tablas de individuos por variables constituyendo lo que se denomina en el contexto estadístico datos de conjuntos múltiples.

### **Análisis Canónico de Correspondencias:**

En general, como las especies son atraídas por condiciones favorables, las variables ambientales son consideradas como explicativas. Por ese motivo, el ACC es un método usual en el que los dos grupos de variables juegan un papel asimétrico. El ACC es un Análisis de Componentes Principales restringido sobre la tabla de especies, que ha sido transformada en una tabla de estadísticos de chi-cuadrado de Pearson. Combina aspectos de ordenación con aspectos del análisis directo del gradiente a través de la regresión. Ter Braak [3] lo presenta como un AC donde los ejes son elegidos en base a las variables ambientales, que participan en forma activa en el análisis. En este caso se selecciona la combinación lineal de las variables ambientales que maximiza la dispersión de las especies.

### **Análisis de Coinercia:**

El ACol estudia la relación entre dos o más conjuntos de variables, basándose en el estadístico de coinercia (Col). Primero se analiza cada tabla por separado. Posteriormente, los resultados de las ordenaciones previas se conectan mediante un análisis de coinercia. La coinercia es una medida global de la coestructura de los sitios en los hiperespacios de las especies y variables ambientales. Es alta cuando las dos estructuras varían simultáneamente, y es baja cuando ellas varían independientemente. El ACol tiene como objetivo hallar el vector  $v_1$  en el espacio de las especies (Y) y un vector  $u_1$  en el espacio de las variables ambientales (X) con máxima coinercia. Cuando Y y X están centradas, entonces el ACol maximiza el cuadrado de las covarianza entre la proyección de X sobre  $u_1$  y la proyección de Y sobre  $v_1$ .

## RESULTADOS Y DISCUSIÓN

El 78,45 % de la inercia total asociada al AC de las frecuencias de especies está explicada por las variables fisicoquímicas. Los dos primeros ejes canónicos del ACC representan el 96,08% y 86% de la inercia proyectada por el mismo eje del AC, indicando que las variables relacionadas con estos factores explican las especies muy satisfactoriamente. La interpretación conjunta de los gráficos (Figuras 1 y 2) permite observar que los sitios de muestreo ubicados en la parte izquierda son sitios más contaminados con altas concentraciones de fósforo, potasio y sulfatos, y asociados a la presencia de especies con mayor tolerancia a la contaminación. Se destacan las especies: *Ephydriidae*, *Glossiphoniidae* y *Tubificidae*. Hacia la parte superior derecha se sitúan los sitios menos contaminados con bajas concentraciones de fósforo, cloro y metales alcalinos con fuerte presencia de especies menos tolerantes a la contaminación (*Aeshnidae*, *Cochliopidae*, *Corydalidae*). El sitio Arroyo Pradera del Puma se separa del resto debido a su gran abundancia de *Simulidea* y *Notonemouridae*. El ACol sustenta lo anterior, indicando que las variables ambientales, especies y sitios proyectados en la misma dirección desde el origen tienen una fuerte asociación. El coeficiente  $RV= 0,46$  indica que existe similitud entre las dos configuraciones, siendo necesario ambos grupos para la caracterización completa de los sitios. En la proyección en el plano factorial de coinercia de los sitios (Figura 2) se visualiza que en los sitios T3 y PP2 hay menor relación entre las especies y variables ambientales. Las dos técnicas

utilizadas conocidas como de Análisis de Gradiente son de ordenación con restricciones, de gradiente. Si bien los roles asumidos por las tablas difieren entre las técnicas, las ordenaciones obtenidas con ambos métodos son similares. El principio en el que se basa el ACol puede ser extendido al caso de más de dos tablas de datos sin la restricción en el número de variables ambientales y de especies en relación con el número de objetos; o en el rol de las diferentes tablas como variables de respuesta o predictoras.

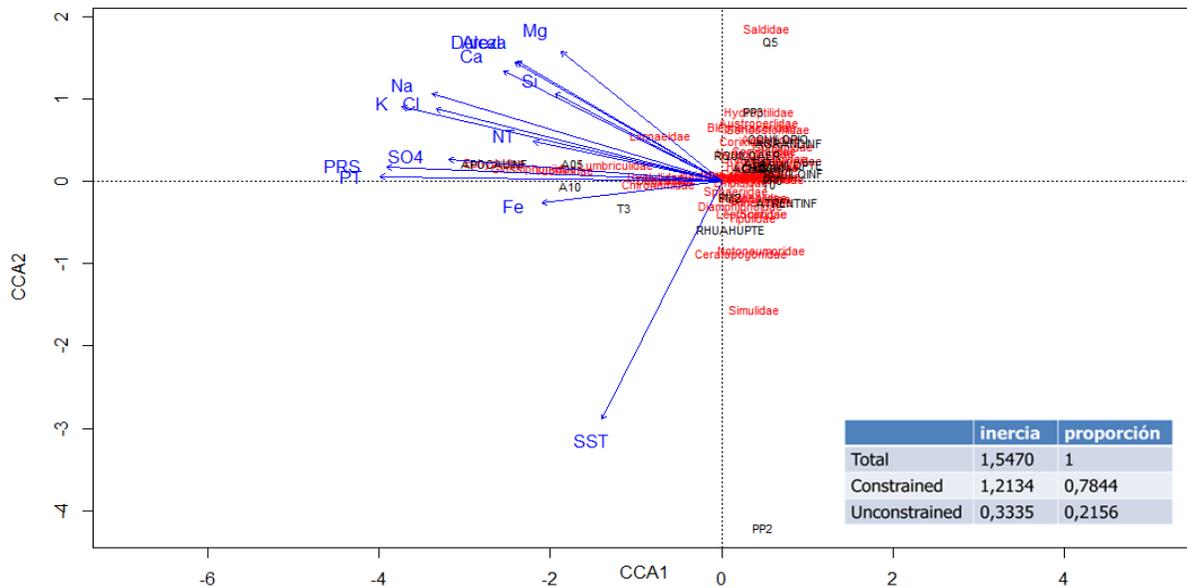


Figura 1: Plano factorial 1 y 2 del ACC:: sitios de muestreo, especies y variables ambientales

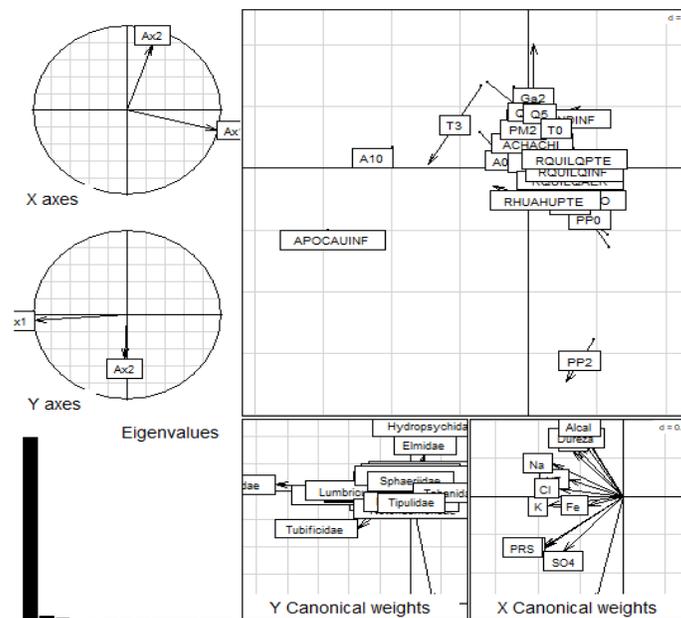


Figura 2: Plano principal del ACol:: sitios de muestreo, especies y variables ambientales



## REFERENCIAS

- [1] Chessel, D. & Mercier, P. (1993). "Couplage de triplets statistiques et liaisons espèces-environment". Publicado en: J.D. Lebretonet B. Asselain, B. (Eds.) *Biometrie et Environnement*. Masson. Paris, pp: 15-44.
- [2] Dolédec, S. & Chessel, D. (1994). "Co-inertia analysis: an alternative method for studying species-environment relationships". *Freshwater Biology*, 31: 277-294.
- [3] Ter Braak, C.F. (1986). "Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis". *Ecology*, 67 (5): 1167-1179.