



## Libro de Resúmenes

9, 10 y 11 de Octubre de 2019  
MENDOZA - ARGENTINA



**UNCUYO**  
UNIVERSIDAD  
NACIONAL DE CUYO

80 años



FACULTAD DE  
**CIENCIAS  
AGRARIAS**



The International  
**Biometric Society**



Grupo Argentino  
de Biometría

XXIV Reunión Científica del Grupo Argentino de Biometría  
XXIV Reunión Científica del Grupo Argentino de Biometría 9,10 y 11 de octubre de  
2019 : Mendoza-Argentina : libro de resúmenes / compilación de Joaquín Llera ... [et al].  
- 1a ed. - Mendoza : Universidad Nacional de Cuyo, 2021.  
Libro digital, PDF

Archivo Digital: descarga y online  
ISBN 978-987-575-218-4

1. Biometría. I. Llera, Joaquín, comp. II. Título.  
CDD 510.72

## Editor

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

## Compiladores

Joaquín A. Llera

Nora B. Martinengo

María Claudia Morelli

María Eugenia Martinez

Eva Marien Maure Russó

ISBN 978-987-575-218-4



La International Biometric Society (IBS) es una sociedad internacional que promueve el desarrollo y la aplicación de la teoría y los métodos estadísticos y matemáticos en las biociencia, 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 IBS llevó a cabo la XXIV Reunión Científica del Grupo Argentino de Biometría (GAB) los días 9, 10 y 11 de Octubre de 2019, organizada por el Grupo Argentino de Biometría y la Facultad de Ciencias Agrarias de la Universidad Nacional de Cuyo (Mendoza).

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, mini-cursos, exposición de trabajos científicos, mesa de análisis de datos, concurso de jóvenes biometristas y la Asamblea Anual Ordinaria del GAB.

## Comisión directiva Región Argentina (GAB)

Presidente: Pablo D. Reeb

Vicepresidente: Silvia Sühling

Secretario: Guillermo Sabino

Tesorero: Maria Inés Urrutia

Vocal Titular: Susana Filippini

Vocal Titular: Gustavo Giménez

Vocal Suplente: María del Carmen Romero

Vocal Suplente: Erika Raña

## Comité Científico

Balzarini, Mónica

Bramardi, Sergio

Casanoves, Fernando

Cendoya, Gabriela

Di Rienzo, Julio

Filippini, Susana

Macchiavelli, Raúl

Reeb, Pablo

Winzer, Nélica

## Comisión Organizadora Local

Joaquín A. Llera

Nora B. Martinengo

María C. Morelli

María E. Martinez

Eva M. Maure Russo

Laura L. Cánovas

Ejes  
BIG DATA  
MODELOS MIXTOS  
MUESTREO



## Contenido

### BIG DATA

Caracterización de Clientes de Tarjetas de crédito de una Institución Financiera usando Análisis Simbólico de Datos.....	2
Aprendizaje automático para la Predicción de Adopción de Mascotas.....	5
Aplicación de algoritmos de Big Data y Modelos No Lineales (MNL) para pronosticar tamaños de frutos a cosecha en peras.....	8

### MODELOS MIXTOS

Modelización de la variabilidad fenotípica de la enfermedad Mal de Río Cuarto en maíz y su asociación con la variabilidad genotípica.....	9
Modelación Bayesiana de la distribución del Carbón de la Espiga del Maíz (Ustilago Maydis) con Modelos Inflados en Cero.....	12
Desigualdades en salud en adolescentes: la escuela como moduladora de la asociación entre nivel socioeconómico y factores de riesgo para sobrepeso y obesidad.	15
Estimación de la estatura en niños hospitalizados a partir de modelos generalizados mixtos.....	16
Influencia de la estructura de varianza-covarianza sobre la evaluación genotípica de amaranto en ensayos multi-ambientales de riego-sequía.....	17
Modelización para la reconstrucción de valores de precipitación diaria.....	18
Modelos Generalizados Mixtos para fenotipado de líneas endocriadas de girasol contrastantes para la tolerancia al estrés hídrico.....	19
Modelos Lineales Generalizados Mixtos para efectos de controles químicos y biológicos en podredumbre ácida en vid.....	20

### MUESTREO

Estimación de la prevalencia de depresión y uso de servicios de salud en San Carlos de Bariloche.....	21
---	----

# Caracterización de Clientes de Tarjetas de crédito de una Institución Financiera usando Análisis Simbólico de Datos

(1) Departamento de Matemática, FFHA, Universidad Nacional de San Juan,

(2) Departamento de Informática, FCEFyN, Universidad Nacional de San Juan. Argentina

MALLEA<sup>(1)</sup>, A., GANGA<sup>(2)</sup>, L.

lamallea@ffha.unsj.edu.ar<sup>(1)</sup>, lamallea@gmail.com<sup>(1)</sup>

## Resumen

Este trabajo se centra en la extracción de conocimiento a partir de fuentes masivas de datos para determinar un perfil que permita describir al usuario de una marca de tarjeta de crédito, con sus diferentes segmentos, de una entidad bancaria. Para ello se utilizan métodos del Análisis Simbólico de Datos ya que no se analiza cada usuario sino clases de usuarios y variables de tipo simbólico, definidas según un objetivo. Se trabaja con dos tablas simbólicas, una donde el concepto es el número de cuenta y otra donde el concepto es el segmento de la tarjeta. Los resultados nos permiten caracterizar a los usuarios de las cuentas de las tarjetas emitidas por la entidad y además encontrar un perfil del usuario según el segmento o producto, de acuerdo a variables de interés tales como: tenedor de plástico, importe de la compra, número de cuotas en que hizo la operación y rubro comercial.

## Palabras Clave

Variable simbólica. Análisis Simbólico de Datos. Caracterización de perfiles.

## Introducción

La ciencia de datos, considerada como una ciencia en sí misma es, en términos generales, la extracción de conocimiento de los datos [7]. Data Mining es una poderosa tecnología con gran potencial para extraer tal conocimiento. Sin embargo, desde el punto de vista estadístico, sus herramientas sólo han sido desarrolladas para trabajar con matrices de datos clásicas, es decir, donde cada unidad es individual y las variables toman un único valor para cada individuo. El análisis de datos simbólicos (SDA, por sus siglas en inglés) [9, 10, 11] brinda una nueva forma de pensar en Data Science al extender la entrada estándar a un conjunto de clases de entidades individuales. Por lo tanto, las clases de una población dada se consideran unidades de una población de nivel superior a estudiar. Tales clases a menudo representan las unidades reales de interés. Para tener en cuenta la variabilidad entre los miembros de cada clase, las mismas se describen por intervalos, distribuciones, conjunto de categorías o números que a veces se ponderan. De esa manera, obtenemos nuevos tipos de datos, llamados "simbólicos", ya que no se pueden reducir a números sin perder mucha información. El primer paso en SDA es construir la tabla de datos simbólicos donde las filas son clases y las variables pueden tomar valores simbólicos. El segundo paso es estudiar y extraer nuevos conocimientos de estos nuevos tipos de datos mediante al menos una extensión de Estadística Computacional y Data Mining a datos simbólicos. SDA es un nuevo paradigma que abre un vasto dominio de investigación y aplicaciones al proporcionar resultados complementarios a los métodos clásicos aplicados a los datos estándar.

SDA también brinda respuestas a los grandes volúmenes de datos (big data) ya que se pueden reducir y resumir por clases.

## Material y Métodos

En este trabajo se presenta una descripción de usuarios de tarjetas de crédito de una entidad financiera, utilizando métodos del análisis simbólico de datos. Se analizan bases de datos referidas a las operaciones con tarjetas de crédito seleccionándose, para tal efecto, la base de datos de clientes de una tarjeta de primera línea (Visa). La misma consta de 623477 registros, que previamente han sido depurados teniendo en cuenta sólo trabajar con consumos en pesos. Por tratarse de datos de una entidad financiera, se han resguardado los mismos. Las variables clásicas seleccionadas para el trabajo son: Edad, Segmento de tarjeta de crédito Visa (Internacional, Gold, Platinum, Black, Signature, Business), Tenedor del plástico (Titular o Adicional), Importe de compra, Cuotas (si realizó el consumo en cuotas) y Rubro comercial.

Estas variables dan origen a variables simbólicas de intervalo, si son de naturaleza cuantitativa, o modales, si son categóricas; cuando en la tabla simbólica de datos el concepto se define como Número de cuenta. Esto es, se ha llevado a cabo una agrupación temporal de la matriz de datos original debido a que en cada cuenta pueden existir varios usuarios que han realizado varias compras con su tarjeta Visa.

A partir de la tabla, así obtenida, se ha realizado una descripción de las variables simbólicas [1, 2, 3], utilizando el módulo DSTAT de SODAS [8] (El software SODAS es el resultado del proyecto ASSO: Analysis System of Symbolic Official Data)

Posteriormente se considera como concepto el Segmento de la tarjeta, resultando una tabla simbólica de seis filas descritas por cinco variables simbólicas. Se ha visualizado la descripción simbólica de estos conceptos con el módulo VIEW de SODAS.

## Resultados

De acuerdo a los resultados obtenidos, la distribución de la variable Importe es muy asimétrica positiva, casi el 98% de las cuentas tienen importes de hasta \$25038 en sus compras, siendo el máximo de \$400617, su media y dispersión de \$13268 y \$6209, respectivamente. La Edad de los tenedores de plástico varía de 13 a 84 años, la media y dispersión son de 49 y 15 años, respectivamente.

En cuanto a las variables simbólicas modales podemos mencionar que hay un 10% de cuentas que tienen tenedores de plástico adicional. El 64,25% de las cuentas corresponde al segmento Visa Internacional, el 26,62% Visa Gold, el 3,62% Visa Nacional, 3,25% Visa Platinum, 1, 26% Visa Signature y el 1% Visa Business.

El 52, 81% de las cuentas registraron consumos en un pago, el 24,19% en tres y el 12,39% en seis pagos, el resto en más cuotas. En cuanto a los rubros, la mayoría de las cuentas hizo operaciones en Supermercado, Indumentaria y Gastos de Seguros, reuniendo un 40% de cuentas entre estos rubros. Se destacan también los gastos en salud, TV y artículos del hogar, con aproximadamente un 6% de cuentas en cada rubro. El resto tiene porcentajes inferiores. Al trabajar con la tabla simbólica donde el concepto es el Segmento (producto) de tarjeta, se observa que existen algunas

diferencias en las cuentas de los seis productos Visa: Visa Business, sólo tiene titulares en sus cuentas, contraponiéndose a Visa Signature que tiene el máximo porcentaje de adicionales entre todos los productos. Los máximos importes se dan en Visa Signature y Visa Platinum y los mínimos en Visa Nacional. Un alto porcentaje de cuentas de todos los productos compra en un pago o tres. Se destaca Visa Business por gastos en el rubro Viajes, los demás productos tienen los mayores porcentajes de gastos en el rubro Supermercado, siendo Visa Nacional e Internacional los máximos de este rubro (20%).

## Conclusiones

Con este trabajo se pretende evidenciar las ventajas del SDA para trabajar clases de individuos de una gran población de interés, sin perder la información de la variabilidad dentro de las clases, como lo haría la estadística clásica al tener que considerar un único valor por unidad (tal como la media para el caso de variables numéricas, o modo para categóricas).

Un futuro trabajo es realizar una clasificación aplicada a la tabla simbólica de cuentas, utilizando Clustering Simbólico Dinámico y Clustering Simbólico Jerárquico. Además se pretende analizar, con métodos del SDA, bases de datos de las diferentes marcas de tarjetas de crédito de la entidad bancaria para poder efectuar comparaciones, a fin de identificar diferencias y/o similitudes entre clases de usuarios de estas marcas.

## Referencias

- [1]Bock, H., Diday, E. (2000). *Analysis of Symbolic Data: Exploratory methods for extracting statistical information from complex data*. Springer-Verlar, Berlin-Heidelberg.
- [2]Billard, L., Diday, E. (2007). *Symbolic Data Analysis: Conceptual Statistics and Data Mining*. Wiley.
- [3]Billard, L., Diday, E. (2003). *From the statistics of data to the statistics of knowledge: Symbolic Data Analysis*. *Journal of the American Statistical Association*, 98 (462).
- [4]Brito, P. (2014). Symbolic Data Analysis: another look at the interaction of Data Mining and Statistics. *WIREs Data Mining and Knowledge Discovery*, Volume 4, Issue 4, July/August 2014, 281-295.
- [5]Calvo, P. (2000). *Aplicaciones de los Objetos Simbólicos en la Estadística Oficial*. Cuaderno Técnico de EUSTAT.
- [6]Calvo P., Prado C., Pérez Y., Ayestarn'a M. (2000). Creación de Objetos Simbólicos a partir de encuestas almacenadas en Bases de Datos Relacionales. *Metodología de Encuestas*, Vol 2, N° 2, 239-260.
- [7]Diday, E (2016) "Thinking by classes in data Science: the symbolic data analysis paradigm". *WIREs Comput Stal*,8:172-205,doi: 0.1002/wics.1384.
- [8]Diday, E. and Noirhomme-Fraiture, M. (2008). *Symbolic Data Analysis and the SODAS Software*. Wiley.
- [9]Diday, E. (2000). *An Introduction to Symbolic Data Analysis and the Sodas Software*. University Paris, Dauphine.
- [10]Diday, E. (2008). *The state of the art in symbolic data analysis: overview and future*.
- [11]Diday, E. (1992). Analyse des données et classification automatique numérique et symbolique. *Seminario Internacional de Estadística en Euskadi*. Volumen 27. EUSTAT.

# Aprendizaje automático para la Predicción de Adopción de Mascotas

DÍAZ.C<sup>1</sup>, FABRO.J<sup>2</sup>

<sup>1</sup> INTA-MANFREDI-Córdoba, Argentina. <sup>2</sup> Pritty, S.A. Córdoba, Argentina,

diaz.carolina@inta.gob.ar

## Resumen

Las tasas de adopción de animales están fuertemente correlacionadas con los metadatos asociados con sus perfiles en línea, como el texto descriptivo y las características de las fotos. Se desarrollarán algoritmos para predecir la adoptabilidad de las mascotas, específicamente, qué tan rápido se adopta una mascota. Se invita a colaborar en la implementación de sus soluciones en las herramientas de IA para evaluar y mejorar el rendimiento de la adopción de mascotas, lo que beneficiará el bienestar animal mundial.

El presente trabajo fue el trabajo premiado de la competencia de adopción de mascotas en donde se aplicaron herramientas de Aprendizaje Supervisado, definimos como métrica para evaluación a la "accuracy" debido a que el objetivo principal era que nos concentremos en los modelos. Dividimos los datos de entrenamiento en "train" y "validation". Se aplicó cross-validation (CV) sólo sobre "train". Luego se utilizó grid search para determinar los hiperparámetros. Se aplicó one-hot encoding para lidiar con las variables categóricas. Para ello, usamos la función "get\_dummies" de pandas. Hicimos esto último sobre todo el universo de datos, asumiendo el riesgo que esto conlleva (puede darse el caso en que creamos muchas nuevas features para valores que no son vistos en los datos de "train"). Se compararon siete algoritmos diferentes arrojando como resultado que el mejor algoritmo predictivo para nuestro dataset es random forest, logrando el mayor puntaje entre los modelos competidores, otorgando al algoritmo un 89% de exactitud para la predicción del problema planteado. Se utilizó al software PYTHON, con la implementación de varios paquetes como Pandas, numpy, sklearn, seaborn.

**Palabras Clave (Cambria, 11, negrita) Hasta cinco.**

Machine Learning, Random Forest, Predicción

## Introducción

Millones de animales callejeros sufren en las calles o son sacrificados en refugios todos los días en todo el mundo. Si se pueden encontrar hogares para ellos, se pueden salvar muchas vidas. PetFinder.my ha sido la plataforma de bienestar animal líder de Malasia desde 2008, con una base de datos de más de 150,000 animales. PetFinder colabora estrechamente con los amantes de los animales, los medios de comunicación, las corporaciones y las organizaciones globales para mejorar el bienestar animal.

En esta competencia, desarrollé algoritmos para predecir la adoptabilidad de las mascotas, se utilizaron herramientas de IA que guiarán a los refugios y rescatistas de todo el mundo a mejorar el atractivo de sus perfiles de mascotas, reduciendo el sufrimiento de los animales y la eutanasia. El trabajo fue premiado en el marco de la competencia de Machine Learning perteneciente a la Diplomatura de Datos de la Facultad de matemática, Astronomía y Física de la UNC, consta de una metodología para la aplicación de herramientas de Aprendizaje Supervisado a partir de la competencia de adopción de mascotas de PetFinder.my en Kaggle: Se ha creado una versión más básica, sin imágenes. Como antecedentes, se pueden mencionar muchos proyectos aplicados a la solución del problema de predicción realizados por los participantes de las competencias Kaggle.

## Material y Métodos

El trabajo consta de encontrar el algoritmo de aprendizaje supervisado con mayor exactitud entre diversos modelos competitivos que mejor realicen la predicción de la rapidez de adopción de mascotas que incluyen a perros y gatos. Entre las features que posee el dataset provisto, se encuentran la raza, la ciudad en donde se encuentran, el color, edad, sexo, edad de maduración, esterilización, salud, calidad y rapidez de adopción. Del dataset se realiza exploración, curación y limpieza de los datos. Se utilizó el 100% del conjunto de "train", teniendo en cuenta que para determinar los hiperparámetros óptimos del algoritmo, utilizaremos cross validations y grid search procedimiento que se encarga de dividir los datos de entrenamiento y validación. El razonamiento se basa en usar todos los datos disponibles para "train" para no desperdiciar información disponible. Las pruebas contra el conjunto de datos de prueba lo realiza kaggle al subir o aplicar el archivo con las predicciones. Aplicamos un conjunto de modelos, que podrían ser utilizados para resolver los objetivos de la competencia, con la mayoría de sus parámetros por defecto de manera de observar cual ajusta mejor. Los algoritmos competidores que utilizamos fueron; Stochastic Gradient Descent (SGD), Logistic Regression, K Nearest Neighbor, Gaussian Naive Bayes, Perceptron, Linear Support Vector Machine, Decision Tree y Random Forest. La mejor exactitud la arroja, Random Forest con parámetros por defecto, obtiene una accuracy promedio de 35,66% +/- 1,1% de desvío estándar. A partir de los resultados realizamos una búsqueda por grilla con el objetivo de encontrar los hiperparámetros que mejoran el rendimiento del modelo. Encontramos los parámetros con mayor performance, Aplicamos nuevamente random forest pero ahora seteando los hiperparámetros encontrados con alta performance. Generamos la predicción del conjunto de validación con el algoritmo random forest obtenido del conjunto de entrenamiento. Por último, el algoritmo fue evaluado por Kaggle con el conjunto total de datos obteniendo un 89% de exactitud para la predicción de la rapidez en la adopción de mascotas.

## Resultados

Como resultado obtuvimos el algoritmo con mayor performance basado en Aprendizaje Supervisado para predecir la rapidez de adopción de mascotas. Random Forest, fue el modelo seleccionado entre ocho modelos competidores, obteniendo el más alto nivel de exactitud que luego de lograr la predicción fue evaluado por el comité de la competición Kaggle EEUU.

Los resultados están presentados en la página competencias Kaggle, en el link <https://www.kaggle.com/c/supervisado2019/leaderboard> Como futuras líneas de trabajo se intentará incorporar imágenes de mascotas para poder aumentar y mejorar la predicción.

## Discusión

Las alternativas al problema planteado recorre un gran abanico de soluciones posibles, que incluyen algoritmos más avanzados y más complejos, sin embargo se necesitaría indagar en la performance de los algoritmos más complejos. La discusión en el uso de los métodos se convierte así en exploración, innovación, y metodologías a investigar.

Al tratarse de una metodología aplicada a un conjunto de datos que no posee la totalidad de las features, se puede considerar al presente trabajo como una aproximación a la predicción objetivo, sin embargo, se presenta como una manera efectiva de la selección de algoritmos basados en Machine Learning.

## Referencias

<https://www.kaggle.com/pierra/credit-card-dataset-svm-classification>

<https://www.kaggle.com/zlatankr/titanic-random-forest-82-78/data>

<https://github.com/susanli2016/Machine-Learning-with-Python/blob/master/Building%20Recommender%20System%20with%20Surprise.ipynb>

# Aplicación de algoritmos de Big Data y Modelos No Lineales (MNL) para pronosticar tamaños de frutos a cosecha en peras

GIMENEZ G.<sup>1</sup>, RUBIO, N., BRAMARDI, S.<sup>1</sup>,

<sup>1</sup>Departamento de Estadística, Facultad de Economía y Administración Universidad Nacional del Comahue

gustavo.gimenez@faea.uncoma.edu.ar

## Palabras claves:

Support Vector Machine, algoritmos de aprendizaje, clasificación multiclase, tamaño de frutos

## Resumen

Tanto el proceso de obtención de conocimientos en bases de datos (KDD) como Big Data, utilizan algoritmos de aprendizaje estadístico. Dichos algoritmos se agrupan en: supervisados y no supervisados. Los algoritmos supervisados para cada observación del predictor requieren un valor de respuesta y se destaca el Support Vector Machine (SVM). El SVM fue concebido como un clasificador binario extendido a problemas de regresión y clasificación multiclase. En este trabajo se aplica el SVM para clasificar frutos en tamaños comerciales desde el diámetro del fruto con 30 a 40 días de antelación a la cosecha. La performance del SVM se comparó con las predicciones de curvas de crecimiento, las cuales se ajustaron utilizando un modelo logístico para cada tamaño comercial a través de MNL. Para ello se analizaron los registros de diámetros en milímetros de peras "Beurre D'Anjou", tomadas 3 semanas posteriores a floración hasta la cosecha, durante 5 temporadas. Los frutos fueron clasificados a cosecha en tamaños comerciales acorde al envase de 18kg. Se seleccionó el 70% de los registros aleatoriamente para entrenar el SVM y ajustar los MNL, el 30% se utilizó para testear ambas técnicas. El algoritmo se entrenó con kernel lineal y el parámetro costo = 16. El testeo se realizó tomando como respuesta el tamaño comercial de cada fruto y predictoras los días de vida y el diámetro del fruto entre los 30 y 40 días antes de cosecha. Para determinar el tamaño comercial mediante MNL, se predice sobre las curvas de crecimiento al momento determinado y se calcula la media de las curvas próximas, si el diámetro del fruto superaba la media se clasificaba con el tamaño superior caso contrario en el tamaño inferior. Para el análisis se utilizó el software R. Los resultados mostraron una precisión del SVM del 60%, superior a los MNL.

# Modelización de la variabilidad fenotípica de la enfermedad Mal de Río Cuarto en maíz y su asociación con la variabilidad genotípica

Rossi Ezequiel<sup>1</sup>, Ruiz Marcos<sup>1</sup>, Balzarini Mónica<sup>2</sup>, Bonamico Natalia<sup>1</sup>

1INIAB-FAV-UNRC, 2CONICET-FCA-UNC, Argentina

E-mail: [erossi@ayv.unrc.edu.ar](mailto:erossi@ayv.unrc.edu.ar)

## Resumen

*El objetivo del trabajo fue modelizar la variabilidad fenotípica de Mal de Río Cuarto (MRC) en líneas diversas de maíz y asociarla con la variabilidad genotípica. Un panel diverso de 185 líneas endocriadas de maíz del CIMMYT se evaluaron durante dos años en dos localidades de la región en donde la enfermedad MRC es endémica. Se evaluaron distintos modelos lineales mixtos para modelar la variabilidad de la incidencia y la severidad de MRC considerando la correlación genética entre las líneas de maíz, la cual se estimó a partir de 86.929 SNPs. El modelo que considera estructura de varianza-covarianza heterogénea para el término de interacción genotipo-ambiente, y la correlación genética entre las líneas fue el que mejor modeló la variabilidad fenotípica de MRC. A partir de esta modelización de los datos fenotípicos, se realizó un mapeo asociativo con la variabilidad genotípica que permitió identificar cuatro SNPs asociados con la incidencia y cinco con la severidad de MRC, a través de ambientes.*

## Palabras Clave

Modelos lineales mixtos, maíz, Mal de Río Cuarto, Incidencia, Severidad.

## Introducción

La investigación en Agricultura comúnmente presenta situaciones donde es necesario utilizar modelos lineales mixtos porque no se cumplen los supuestos de independencia, normalidad o igualdad de varianzas. En estudios donde se evalúan distintos genotipos en varios ambientes es razonable asumir que existen correlaciones entre observaciones del mismo ambiente (Casanoves, 2004). La modelización de la covarianza genotipo-ambiente permite obtener conclusiones más confiables sobre las diferencias fenotípicas entre los genotipos (Malosetti et al., 2008). El objetivo del trabajo fue modelizar la variabilidad fenotípica de Mal de Río Cuarto (MRC) y asociarla con la variabilidad genotípica de líneas diversas de maíz.

## Material y Métodos

Material vegetal y ensayo de campo.

El material vegetal está constituido por un panel de 185 líneas endocriadas de maíz del CIMMYT. La incidencia (INC-MRC) y la severidad (SEV-MRC) de MRC se evaluaron durante 2017-18 y 2018-19 en cuatro ambientes del sur de Córdoba, Argentina: Sampacho 2017-18 (SA18) y 2018-19 (SA19), Río Cuarto 2017-18 (RC18) y 2018-19 (RC19). En cada ambiente se utilizó un diseño parcialmente repetido (Williams et al., 2011) con un 25% de los genotipos con tres repeticiones. Datos fenotípicos. Cada planta se clasificó por el grado de severidad de enfermedad según la escala propuesta por Ornaghi et al., (1999). El carácter INC-MRC es el porcentaje de plantas con síntomas respecto al total de plantas de cada unidad experimental. La SEV-MRC es el grado medio de las plantas con síntomas.

Datos genotípicos.

La caracterización de las líneas de maíz de CIMMYT con SNPs es de disponibilidad pública (Wu et al. 2016). A partir de 362.008 SNPs, se seleccionaron 86.929 en los 10 cromosomas, considerando una tasa de datos faltantes menor al 50% y una frecuencia superior al 0,05 en el alelo menos frecuente. Análisis estadístico. Los datos fenotípicos se analizaron con un modelo mixto multivariado, paquete sommer (Covarrubias-Pazarán, 2016), software R (R Core Team 2016). Para modelar la covarianza genotipo-ambiente, distintas estructuras de varianza-covarianza fueron evaluadas (homogénea, heterogénea y no estructurada). El mejor modelo se definió mediante la prueba del cociente de verosimilitud. A su vez, cada modelo se probó también considerando la correlación genética entre las líneas. Para comparar modelos con la misma estructura de varianza-covarianza considerando o no la correlación genética, se utilizó el criterio de Akaike (AIC). El modelo mixto multivariado permitió estimar la correlación entre los caracteres INC-MRC y SEV-MRC. Posteriormente, un modelo mixto se utilizó para realizar la asociación entre la variabilidad fenotípica (INC-MRC y SEV-MRC) y la variabilidad genotípica (86.929 SNPs). La matriz de relación aditiva fue estimada con la función A.mat del paquete sommer (Covarrubias-Pazarán, 2016), software R (R Core Team 2016). La corrección por multiplicidad propuesta por Li y Ji (2005) fue usada para controlar falsos positivos.

## Resultados y Discusión

El modelo que considera estructura de varianza-covarianza heterogénea ( $p < 0,0001$ ) y correlación genética entre las líneas fue el que mejor ajustó la variabilidad de los caracteres incidencia y severidad de Mal de Río Cuarto. Esto indica que cada ambiente tiene su propia varianza genética y no hay correlación entre los ambientes. Esto es esperable dado que la enfermedad Mal de Río Cuarto es transmitida por insectos, y por lo tanto la presión natural de la enfermedad varía con los ambientes. Tal como plantea Malosetti et al., (2008) la modelización de la covarianza genotipo-ambiente permite obtener conclusiones más confiables sobre las diferencias fenotípicas entre los genotipos. La correlación entre ambos caracteres a través de ambientes fue estadísticamente significativa con un valor de 0,58. Además, se observó amplia variabilidad fenotípica tanto para incidencia como para severidad. La asociación de la variabilidad fenotípica

con la variabilidad genotípica permitió identificar cuatro y cinco SNPs asociados significativamente con incidencia y severidad de Mal de Río Cuarto, respectivamente. Investigaciones adicionales son requeridas para validar los resultados obtenidos.

## Conclusiones

La modelización de la variabilidad fenotípica del MRC en maíz y su asociación con la variabilidad genotípica permite una identificación precisa de regiones genómicas para resistencia a esta enfermedad.

## Referencias

- Casanoves, F. 2004. Análisis de ensayos comparativos de rendimiento en mejoramiento vegetal en el marco de los modelos lineales mixtos. Tesis de doctorado, Universidad Nacional de Córdoba. 143p
- Covarrubias-Pazaran, G. 2016. Genome-assisted prediction of quantitative traits using the R package sommer. PLoS ONE 11(6):E0156744.
- Li J, L Ji. 2005. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. Heredity. 95:221–227.
- Malosetti M, JM Ribaut, M Vargas, J Crossa, FA van Eeuwijk. 2008. A multi-trait multi-environment QTL mixed model with an application to drought and nitrogen stress trials in maize (*Zea mays* L.). Euphytica 161:241–257.
- Ornaghi JA, GJ March, GT Boito, A Marinelli, JE Beviacqua, J Giuggia, SL Lenardon. 1999. Infectivity in natural populations of Delphacodes kuscheli vector of “Mal Río Cuarto” Virus. Maydica 44: 219-223.
- R Core Team. 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, URL (<http://www.R-project.org/>).
- Williams E, HP Piepho, D Whitaker. 2011. Augmented p-rep designs. Biometrical Journal, 53:19–27.
- Wu, Y, F San Vicente, K Huang, T Dhliwayo, DE Costich, K Semagn, N Sudha, M Olsen, B Prasanna, X Zhang, R Babu. (2016). Molecular characterization of CIMMYT maize inbred lines with genotyping-by-sequencing SNPs. Theoretical and Applied Genetics, 129:753–765.

# Modelación Bayesiana de la distribución del Carbón de la Espiga del Maíz (*Ustilago Maydis*) con Modelos Inflados en Cero

Videla María Eugenia<sup>1,2,4</sup>, Kistner Belén<sup>1,3</sup>, Iglesias Juliana<sup>1,3</sup> y 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 vinculado a UFyMA (Unidad de Fitopatología y Modelización Agrícola INTA -CONICET).

<sup>3</sup> Mejoramiento Genético de Maíz. EEA INTA Pergamino.

<sup>4</sup> Universidad Nacional de Villa María (UNVM)

eugeniavidela12@gmail.com

## Resumen

*Algunas variables epidemiológicas en cultivos agrícolas se miden como cantidad de plantas enfermas en una parcela (conteos) y tienen la particularidad de contener una gran proporción de valores iguales a cero. El objetivo de este trabajo es comparar el desempeño de diferentes modelos bayesianos ajustados sobre variables de conteo utilizando el método de estimación por aproximación de Laplace anidada integrada (INLA- Integrated Nested Laplace Approximation). Se ilustra sobre un conjunto de 79 líneas estables de maíz con carbón de la espiga. Los modelos evaluados fueron MLGM Poisson, MLGM Binomial Negativo, Modelo Inflado en Cero Poisson y Modelo Inflado en Cero Binomial Negativo.*

**Palabras Clave: Modelos Lineales Generalizados Mixtos – Variables Discretas – Conteos - INLA**

## Introducción

Las variables epidemiológicas que afectan a cultivos agrícolas suelen registrarse con variables discretas (plantas enfermas de una parcela) y modelarse a través de Modelos Lineales Generalizados (MLG) (Diaz-Quijano, 2016). En estudios para identificar genotipos resistentes al carbón de la espiga del maíz (*Ustilago Maydis*) la opción clásica es ajustar Modelos Lineales Generalizados Mixtos (MLGM) con distribución Poisson. Sin embargo, estos modelos son vulnerables a la sobredispersión debido a la relación de igualdad que existe entre media y varianza. En este contexto, los MLGM basados en la distribución Binomial Negativa incorporan un componente de variabilidad para datos con sobredispersión (Ismail y Jemain, 2007). Otro desafío en este tipo de variables es que contienen una gran proporción de valores iguales a cero (plantas sin síntoma en toda la parcela de estudio), lo cual también genera sobredispersión (Phang y Loh, 2013). Estos ceros pueden clasificarse en “falsos ceros”, causados por parcelas que no estén expuestas a la enfermedad o por errores del observador y en “ceros auténticos” que son aquellos generados por parcelas que han estado expuesta al hongo pero por diversos motivos (resistencia del genotipo, condiciones ambientales, etc) ninguna planta ha sido infectada. Distinguir unos ceros de otros es fundamental para identificar genotipos resistentes. Los MLGM de tipo “Inflados en Cero” (MIC), surgen como una alternativa para ajustar el exceso de ceros y evitar la sobredispersión. Para cada observación se producen dos procesos, el primer proceso genera el dato para cada observación como resultado de un ensayo Bernoulli con probabilidad, asumiendo que existen dos posibles resultados para cada observación (sana/enferma), mientras que el segundo

proceso genera los resultados de una distribución paramétrica específica, como Poisson (Modelo MICP) o binomial negativo (Modelo MICBN) con probabilidad (Lambert, 1992). Recientemente, se ha desarrollado un método de estimación para ajustar modelos bayesianos bajo la aproximación de Laplace, INLA por sus siglas en inglés Integrated Nested Laplace Approximation (Rue et al., 2009) como una alternativa computacionalmente eficiente para el método de cadenas de Markov Monte Carlo (MCMC) (Blangiardo et al., 2013).

El objetivo de este trabajo es comparar el desempeño de modelos bayesianos, ajustados sobre variables de conteo, que no contemplan los excesos de ceros (MLGM Poisson (P) y MLGM Binomial Negativa (BN)) respecto a modelos de conteo para datos inflados en cero (MICP y MICBN) para modelar la distribución del carbón de la espiga del maíz con la finalidad de seleccionar genotipos resistentes.

## Material y Métodos

Se evaluaron 79 genotipos de maíz pertenecientes al grupo de mejoramiento de maíz de INTA Pergamino, Argentina en dos localidades (Tucumán y Pergamino). Se contabilizaron la cantidad de plantas enfermas con carbón de la espiga del maíz (*Ustilago maydis*) sobre parcelas con 30 plantas. El 50% de los datos fueron 0. Se ajustaron cuatro modelos bayesianos para datos de conteo: MLGM Poisson (P), MLGM Binomial Negativa (BN), Modelo Poisson Inflado en Cero (MICP) y Modelo Binomial Negativa Inflado en Cero (MICBN). Los modelos se ajustaron utilizando el método INLA implementado en la función `inla()` de la librería INLA de R (Martino y Rue, 2010). Los modelos contemplaron el efecto de localidad y de repetición dentro de localidad como fijo y el efecto de genotipo como aleatorio para estimar el mejor predictor lineal insesgado (BLUP) y posteriormente realizar un ranking de los genotipos de acuerdo a su comportamiento de resistencia al carbón de la espiga. El desempeño de los modelos se comparó con el criterio de información de deviance (DIC), número de parámetros efectivos ( $p$ ) y marginal log-Likelihood (MlogLink).

## Resultados

*Tabla 1: Criterio de información de Deviance (DIC), número de parámetros efectivos ( $p$ ) y marginal log-Likelihood (MlogLink) obtenidos al ajustar 4 modelos.*

Modelo	DIC	$p$	MlogLink
P	809.706	62.16	-464.40
BN	789.692	58.34	-445.65
MICP	794.713	60.27	-455.96
MICBN	788.346	59.34	-446.24

*P: MLGM Poisson, NB: MLGM Binomial Negativa, MICP: Modelo Inflado en Cero Poisson, MICBN: Modelo Inflado en Cero Binomial Negativa.*

Tabla 2: Valores promedios de los efectos aleatorios de cada genotipo para los más susceptibles (valores mayores a 2) y los menos susceptibles (valores menores a -1) obtenidos con el Modelo MICBN.

Genotipo	BLUP
L4674	3.290987
LP605	2.882328
LP915	2.734550
LP579	2.407617
LP1512	-1.142281

Los modelos con distribución Binomial Negativa (BN y MICBN) tuvieron menos cantidad de parámetros efectivos ( $p$ ) que los Poisson (P y MICP) y mayor probabilidad marginal (MlogLink). El modelo inflado en cero Binomial Negativo (MICBN) tuvo mejor ajuste (menor DIC) que el modelo inflado en cero Poisson (MICP) y que los modelos de conteo tradicionales (P y BN) (Tabla 1).

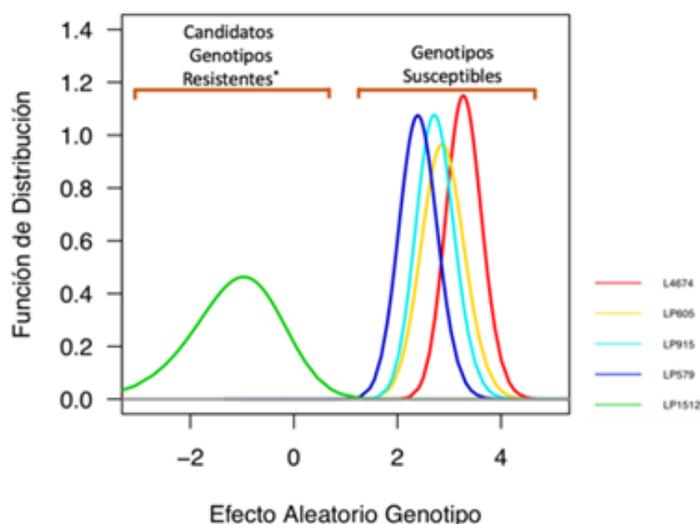


Figura 1: Densidades de los efectos aleatorios de los genotipos más susceptible y más resistentes al carbón del maíz.

\*La densidad de los 18 genotipos más resistentes fueron idénticas. Se muestra sólo para el LP1512.

Con el MICBN se realizó un ranking del efecto aleatorio del genotipo. Se destacaron cuatro genotipos con valores promedios de efecto aleatorio de genotipo mayor a 2 (genotipos más susceptibles) y 18 genotipos con valores promedios de efecto aleatorio menores a -1 (más resistentes) según el modelo seleccionado. Los genotipos seleccionados como más resistentes presentaron idéntico valor medio y sus densidades se superpusieron, por tal motivo solo se

muestra la función de distribución del genotipo LP1512 como representante de este grupo de genotipos (Tabla 2 – Figura 1).

### Conclusión

El método INLA permitió obtener predictores del merito genético en ensayos a campo para selección de genotipos con excesos de cero para una variable discreta. Los modelos Inflados en Cero con distribución Binomial Negativo resultaron más precisos con respecto a los modelos tradicionales como MLGM Poisson y Binomial Negativo o el modelo de Poisson Inflado en Cero.

### Referencias

- Blangiardo, M., Cameletti, M., Baio, G., y Rue, H. (2013). Spatial and spatio-temporal models with R-INLA. *Spatial and spatio-temporal epidemiology*, 4, 33-49.
- Diaz-Quijano, F. (2016). Regressions applied to the study of discrete events in epidemiology. *Revista de la Universidad Industrial de Santander. Salud*, 48(1), 9-15.
- Ismail, N., y Jemain, A. (2007). Handling overdispersion with negative binomial and generalized Poisson regression models. *In Casualty Actuarial Society Forum* (pp. 103-158).
- Lambert, D. (1992). Zero-inflated Poisson regression models with an application to defects in manufacturing. *Technometrics* 34(1), 1 – 14.
- Martino, S. y Rue, H. (2010). Implementing approximate bayesian inference using integrated nested laplace approximation: a manual for the inla program. Available from: <http://www.math.ntnu.no/hrue/>
- Phang, Y. y Loh, E. (2013). Zero inflated models for overdispersed count data. *Int. J. Math. Comput. Nat. Phys. Eng*, 7, 817-819.
- Rue, H., Martino, S. y Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J R Stat Soc B*;71(2):1–35.

## **Desigualdades en salud en adolescentes: la escuela como moduladora de la asociación entre nivel socioeconómico y factores de riesgo para sobrepeso y obesidad.**

ARNAUDO M., FERNÁNDEZ M., KONFINO, J. y PÉREZ A.,

Grupo de Bioestadística Aplicada, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires

beluarnaud@gmail.com

### **Palabras Clave**

Encuesta Mundial de Salud Escolar, análisis multinivel, factores de riesgo y protección

### **Resumen**

El sobrepeso está aumentando a nivel mundial. En Argentina, 3 de cada 10 niños o adolescentes tienen exceso de peso, siendo afectados en mayor medida aquellos con nivel socioeconómico (NSE) más bajo. A fin de analizar cómo la prevalencia de factores de riesgo y de protección que se sabe que se relacionan con la epidemia varían según NSE a distintas escalas se trabajó con información recolectada en la Encuesta Mundial de Salud Escolar (EMSE, 2012). Participaron 21745 estudiantes de 1ro a 3er año de 561 escuelas secundarias del país. Para modelar el consumo de frutas, de bebidas azucaradas y el nivel de actividad física se aplicaron modelos lineales generalizados mixtos con distribución bernoulli. Se incluyeron como variables predictoras el nivel educativo máximo alcanzado por los padres, el nivel educativo promedio de la escuela y el porcentaje de hogares con necesidades básicas insatisfechas (NBI) de la provincia, todas ellas como proxies de NSE a distintas escalas. Los modelos fueron ajustados por sexo y edad, y se incluyó a la escuela como factor aleatorio. Para las tres variables respuesta se encontró interacción entre el NSE del hogar y de la escuela, mientras que el NBI de la provincia resultó no significativo. La prevalencia de consumo de frutas y de alto nivel de actividad física aumenta con el NSE del hogar. Sin embargo, los estudiantes de hogares de menores recursos que asisten a escuelas de mayor NSE tienen mayor prevalencia que los que asisten a escuelas de menor NSE. El consumo de bebidas azucaradas aumenta con el NSE del hogar, pero en este caso estudiantes de bajos recursos que concurren a escuelas de mayor NSE tienen mayor prevalencia que los que asisten a escuelas de menor NSE. Aplicar políticas a escala escuela puede modificar las predicciones para todos los estudiantes, beneficiando principalmente a aquellos provenientes de hogares de menores recursos.

## Estimación de la estatura en niños hospitalizados a partir de modelos generalizados mixtos

Filippini, O. S<sup>1</sup>; Martinez, C.<sup>1</sup>; Ibero, A<sup>1</sup>

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

sfilippini@unlu.edu.ar

Palabras Clave: Estatura, niños, Modelo lineal Generalizado Mixto, Precisión.

### Resumen

La medición de la estatura es una de las técnicas antropométricas que más se utiliza como indicador del crecimiento y estado nutricional de niños en el análisis clínico. Con ella se pueden construir indicadores como el Índice de Masa Corporal y otros. En muchas ocasiones no es posible medir la estatura de niños hospitalizados. Según lo informado en alguna oportunidad por el servicio de crecimiento y Desarrollo del Hospital de pediatría Prof. Dr. J. P. Garrahan, en un 24 % de niños internados no fue posible medir la estatura por múltiples causas, como dolor, inmovilización, tubos de drenajes, etc. Tener mediciones con máxima precisión y mínimo error es indispensable cuando se intenta evaluar el crecimiento en función de la estatura. El objetivo de este trabajo es identificar un modelo para predecir la estatura en la niñez que permita mediante la medición sencilla de distintos miembros, precisión y reproducibilidad. A tal fin se considera una muestra de 210 niños del hospital mencionado, donde se analizan las siguientes variables: Sexo, edad, antebrazo, tibia, Tanner, pierna, peso, cúbito y estatura. Se comparan diferentes modelos lineales. Se selecciona un modelo Lineal Generalizado Mixto de la Estatura en función de las diferentes mediciones de segmentos corporales. El Modelo desarrollado por Gauld sobre el que se validan los datos argentinos propone utilizar el Cúbito y la edad como variables simples de medir en cualquier caso. La función seleccionada con la muestra utilizada en función de estas dos características, presenta un 94% de variación explicada.

# Influencia de la estructura de varianza-covarianza sobre la evaluación genotípica de amaranto en ensayos multi-ambientales de riego-sequía

SÜHRING S.<sup>1,2</sup>, PIEPHO H.<sup>3</sup>, ANDRADE A.<sup>4</sup>, VELÁSQUEZ B.<sup>4</sup>, CURTI R.<sup>1,2</sup>

<sup>1</sup> Cátedra de Estadística y Diseño Experimental, Facultad de Ciencias Naturales, Universidad Nacional de Salta, Av. Bolivia 5150, (4400) Salta, Argentina.

<sup>2</sup> Laboratorio de Investigaciones Botánicas (LABIBO), Facultad de Ciencias Naturales, Universidad Nacional de Salta, Argentina.

<sup>3</sup> Biostatistics Unit, Institute of Crop Science, University of Hohenheim, Stuttgart, Germany

<sup>4</sup> Universidad Nacional de Jujuy, Argentina

e-mail: ssuhring@gmail.com

**Palabras clave:** *Amaranthus spp.*, Ensayo comparativo de variedades, Heterogeneidad de varianzas, Modelos lineales mixtos.

## Resumen

En los programas de mejoramiento de cultivos las recomendaciones varietales se realizan en función del valor genotípico. El valor genotípico se estima por medio de ensayos comparativos de rendimiento en múltiples ambientes (ECR). En los ECR es común encontrar varianzas-covarianzas heterogéneas debido a diferencias genotípicas y/o ambientales, por lo que los análisis que utilizan modelos lineales tradicionales (ANOVA), raramente ofrecen estimaciones precisas de los rendimientos promedios. En este trabajo exploramos la influencia de la estructura de varianza-covarianza sobre la valoración genotípica de amaranto en ensayos multi-ambientales de riego-sequía. El rendimiento de grano de 11 genotipos de amaranto andino fue evaluado en ensayos pareados de riego-sequía en cuatro localidades de Jujuy. Para cada ensayo (control vs sequía) se siguió un diseño en bloques completos al azar con cuatro repeticiones. Los modelos lineales mixtos propuestos para evaluar el rendimiento de los genotipos correspondieron a tres estructuras de varianza-covarianza (componentes de varianza, simetría compuesta y desestructurada), considerando la falta de restricción a la aleatorización producto del arreglo de los tratamientos en cada localidad e incluyendo o no la heterogeneidad de varianzas por tratamiento (riego). Para la selección del modelo más adecuado se utilizó el criterio de información de Akaike. Además, se evaluó la consistencia en los resultados de las pruebas de comparaciones de medias estimadas por cada modelo. El modelo que consideró varianza-covarianza desestructurada y varianzas distintas para tratamientos resultó el más adecuado en términos del valor AIC. Sin embargo, la diferenciación entre genotipos para el rendimiento en grano fue consistente entre los distintos modelos (valores de índice de consistencia entre 97 y 100%) y las estimaciones BLUE de esos rendimientos promedio fueron muy semejantes entre los modelos (cerca de un 5% de variabilidad).

## Modelización para la reconstrucción de valores de precipitación diaria

Paola Luciana Salvatierra<sup>1,2</sup>, María Eugenia Videla<sup>1,2,4,5</sup> y Fernando Forgioni<sup>2,3</sup>

<sup>1</sup> Cátedra de Bioestadística. Instituto de Ciencias Básicas y Aplicadas (IAPCByA). Universidad Nacional de Villa María (UNVM).

<sup>2</sup> Observatorio Regional de Cambio Climático (ORCC). Universidad Nacional de Villa María.

<sup>3</sup> Cátedra de Climatología Agropecuaria. Instituto de Ciencias Básicas y Aplicadas (IAPCByA). Universidad Nacional de Villa María (UNVM).

<sup>4</sup> Cátedra de Estadística y Biometría. Facultad de Ciencias Agropecuarias (FCA). Universidad Nacional de Córdoba. Grupo vinculado a UFyMA (Unidad de Fitopatología y Modelización Agrícola INTA -CONICET).

<sup>5</sup> Consejo Nacional de Investigaciones Científicas y Tecnológicas (CONICET)

salvatierrapaolaluciana@gmail.com

**Palabras Clave:** Serie de Tiempo – Modelo Lineal Generalizado Mixto – Modelo Inflado en Cero

### Resumen

La precipitación diaria ha sido relegada en la investigación del clima a un segundo plano por la dificultad que presentan las series diarias ya que un gran porcentaje de las observaciones corresponden a 0 (no hay precipitación), lo cual genera sobredispersión en los datos que dificulta la modelización a través de los Modelos Lineales Generalizados Mixtos clásicos. En este trabajo se propone la modelización de una serie completa de Enero de 2016 a Julio de 2019 de la Estación Meteorológica del Campo Experimental de la Universidad Nacional de Villa María con el objetivo de reconstruir series de precipitación diaria que permita completar datos faltantes. Con un modelo inflado en cero para variables continuas se calculó, para cada día, dos valores de referencia: una predicción binomial que expresa la probabilidad de que en un día haya o no haya precipitación y una predicción de magnitud que estima la cantidad de precipitación. Para calcular estos valores de referencia se utilizan datos de Temperatura Mínima (°C) y Humedad (%HR) diaria. Para la validación del modelo se estimó el error medio de predicción a través de el procedimiento jackknife.

## Modelos Generalizados Mixtos para fenotipado de líneas endocriadas de girasol contrastantes para la tolerancia al estrés hídrico

DIAZ, C<sup>1</sup>, GRANDON, N<sup>1</sup>, MARTIN, E<sup>2</sup>, CICONI, E<sup>3</sup>, MAMANÍ, E<sup>1</sup>, YACHETTA, G<sup>4</sup>, CASALANGUIDA, M<sup>4</sup>, MECCHIA, L<sup>4</sup>, MORENO, M<sup>1</sup>.

<sup>1</sup> INTA-EEA Manfredi. Ruta Nac. N° 9. Km 636. (5988) Manfredi, Córdoba, Argentina

<sup>2</sup> IICAR-CONICET. Campo Experimental Villarino, Zavalla, Santa Fe Argentina. <sup>3</sup> Asesor técnico.

<sup>4</sup> Instituto A. P. Ciencias Básicas y Aplicadas. Universidad Nacional de Villa María, Córdoba, Argentina.

diaz.carolina@inta.gob.ar

### Palabras Clave

Glimmix, Beta, Fenotipado, Girasol, Ecofisiología.

### Resumen

El objetivo fue caracterizar siete líneas endogámicas de girasol (IL) (*Helianthus annuus* var. macrocarpus (DC) Cockerell) para la tolerancia a la sequía en condiciones controladas. Se realizó Fenotipación en invernadero durante 16 días en etapa vegetativa bajo tratamientos regados (CR) y con déficit de agua (SR). Las variables ecofisiológicas observadas: ganancia de área foliar (GAF), uso total de agua (UAT), tasa de asimilación neta (TAN), eficiencia del uso del agua (EUA) y respuesta de la tasa de transpiración (TT) al déficit de presión de vapor (DPV) (pendiente y punto de ruptura), área foliar inicial (AFi). Se conduce un diseño en bloques incompletos "alpha látice", para controlar variabilidades espaciales, considerando 2 factores; líneas de girasol (7 niveles) y riego (2 niveles) que conforman un total de 14 tratamientos, con 5 repeticiones. Se utilizaron modelos lineales generalizados mixtos, distribución beta, enlace logit, efectos fijos tratamientos, efectos aleatorios columna y fila. Regresiones logísticas asumiendo área foliar inicial (AFi) como covariable. Se encontraron diferencias significativas entre genotipos y riego. Se observó una interacción significativa para TAN, EUA y punto de interrupción en la respuesta DPV. En condiciones sin riego (SR), los genotipos mostraron niveles de GAF y UAT más bajos, en la misma condición TAN y EUA aumentaron y mostraron una alta correlación entre ellos ( $r = 0.77$ ,  $p < 0.0001$ ). Todos los genotipos mostraron reducción de la pendiente y el punto de ruptura en la alta respuesta del DPV en condiciones sin riego (SR). El PCA explicó 80% de variación fenotípica. PC1 discriminó HA89 y R419 mientras que PC2 discriminó por EUA y UTA. Según estos resultados, las IL se clasificaron en alta (HA64 y HAR4), intermedia (B59, R423 y R432) y baja (HA89 y R419). Software, SAS UE, versión 2018., PROC GLIMMIX.

## Modelos Lineales Generalizados Mixtos para efectos de controles químicos y biológicos en podredumbre ácida en vid

DÍAZ, C. D.P.<sup>1</sup>, CORDES, D.<sup>2</sup>

<sup>1</sup> INTA-EEA Manfredi. Ruta Nac. N° 9. Km 636. (5988) Manfredi, Córdoba, Argentina.

<sup>2</sup> INTA-Agencia de Extensión Jesús María, Córdoba, Argentina.

[diaz.carolina@inta.gob.ar](mailto:diaz.carolina@inta.gob.ar)

### Palabras Clave.

Generalizados, Mixtos, Incidencia, Vid, Podredumbre

### Resumen

La podredumbre ácida, enfermedad producida por complejo de microorganismos; levaduras, hongos y bacterias, afectan el rendimiento de vid, con pérdidas mayores al 50% en viñedos de Colonia Caroya, Córdoba, impactando negativamente la economía de los productores. El presente trabajo evaluó el efecto de estrategias de control para disminución de daños causados por podredumbre en vid. El ensayo se realizó bajo DBCA en 4 bloques, formados por un espaldero de vid, 6 plantas adultas de la variedad Ancelotta, UE evaluada durante tres campañas (años 2015/2016/, 2016/2017, 2017/2018). Se dispusieron 4 tratamientos (combinación de productos químicos y fenología) determinados a priori por experiencia de productores. T1: Testigo; T2: Azoxitrobina Zoxitrobina, Difenconazole en floración, T3 Ciprodinil, Fludioxon en cierre de racimo y envero, Aceite del té en madurez, Boscalid, Piraclostrobin en Floración y cierre racimo, Tebucanazole, Trifloxistrobin envero, Procimidone en madurez y T4 Trichoderma 5 aplicaciones desde floración cada 15 días. . Se evaluó, Incidencia (N° de racimos dañados por podredumbre ácida /total de racimos) y Severidad expresada en % del rendimiento, dañado. Se utilizaron MLGM, distribución binomial y norma , covtest para selección de efectos fijos; año, tratamiento y año\* tratamiento, efectos aleatorios bloques\*año, matriz de Var- Cov con AR(1). LSM, Alpha=0.05 para comparación de medias, Tukey-Kramer para comparación multiple. Estimation Technique Residual PL, y Reml. Criterios de bondad de ajuste, -2 Res Log Pseudo-Likelihood, Gener. Chi-Square / DF y AIC, BIC para normal. Se observaron diferencias significativas entre tratamientos. En Incidencia, la mínima estimación 5,7% para T3 y la máxima estimación 48,9% para testigo. Para severidad en T3 se observó 5,89% mientras que testigo arrojó 41,2%. Para T2 se observaron (Inc. 16,5% y Sev. 16,4%.) diferencias significativas con respecto T3 y T1. No existieron diferencias significativas entre T4 y T1. Software SAS UE, PROC GLIMMIX.

## Estimación de la prevalencia de depresión y uso de servicios de salud en San Carlos de Bariloche

HUAYLLA, C.<sup>1</sup>, ZACHARIAS, D.<sup>2</sup>, GARIBOTTI G.<sup>2</sup>

<sup>1</sup> Instituto de Investigaciones en Recursos Naturales, Agroecología y Desarrollo Rural (IRNAD), Sede Andina, Universidad Nacional de Río Negro, Bariloche. CONICET.

<sup>2</sup> Departamento de Estadística, Centro Regional Universitario Bariloche, UN Comahue.

cahuaylla@unrn.edu.ar

**Palabras clave:** Depresión, Muestreo Estratificado, Radio censal.

### Resumen

**Introducción:** La depresión es una enfermedad mental y según la OMS es la principal causa de discapacidad a nivel mundial.

**Metodología:** Se realizó una encuesta cara a cara en hogares a una muestra estratificada, polietápica y probabilística de 300 personas de 18 a 64 años residentes en la ciudad. El instrumento de recolección de datos incluyó la escala CESD-R. Etapas de muestreo: i) selección de radios censales siguiendo un muestreo estratificado por quintil de un índice de nivel socioeconómico especialmente diseñado en base a los datos del censo de 2010, considerando el promedio de años de estudio de los miembros del hogar mayores de 25 años. En cada quintil se realizó un muestreo aleatorio simple de radios censales, ii) selección sistemática con inicio aleatorio del hogar; iii) selección de la persona a entrevistar: muestreo de cuotas según género y grupo etario. Se realizó la prueba chi-cuadrado para evaluar diferencias en la prevalencia de depresión subclínica.

**Resultados:** La media de edad fue 39,3 años y el desvío estándar 13,6 años. En relación al nivel educativo, 6% no completó ningún nivel, 29% completó el nivel primario, 44% el nivel secundario y 21% estudios terciarios/universitarios. La distribución por género, edad y nivel educativo fue muy similar a la de la población de Bariloche de 18 a 64 años. La prevalencia de depresión mayor fue 5,7% (IC95%: 3,3-8,9), 7,2% en estratos de menor nivel socioeconómico y 3,4% en los de mayor nivel socioeconómico, de los cuales buscaron tratamiento el 38,5% y el 60%, respectivamente.

**Conclusiones:** La prevalencia de depresión mayor en Bariloche es similar a la observada a nivel nacional. El estudio destaca una brecha importante en el uso del servicio y el tratamiento de la depresión, siendo mayor en los estratos socioeconómicos más bajos donde es más alta también la prevalencia de depresión.